Queen Mary University of London
School of Biological and Chemical Sciences

Structural studies of the *Dickeya dadantii* type II secretion system protein GspB

A THESIS SUBMITTED BY

Piers Rycroft

IN PARTIAL FULFILMENT OF THE REQUIREMENT OF THE
UNIVERSITY OF LONDON
FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

AUGUST 2015

Declaration

I, Piers Benjamin Rycroft, confirm that the research included within this thesis is my own work or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure that the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept that the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature: *Piers B. Rycroft*

Date: 8th September 2015

Acknowledgements

Supervisor:	Prof. Richard Pickersgill	(QMUL)
Supervisory panel members:	Dr John Viles	(QMUL)
	Prof. Conrad Mullineux	(QMUL)
Collaborators	Dr Vladimir Shevchik	(Université Lyon)
	Dr Geoff Kelly	(NIMR NMR Centre)
Colleagues	Miss Yumiko Tashiro	
	Dr Shuang Gu	
	Dr Saima Rehman	
	Dr Allan Pang	
	Dr Nadine Younan	
	Dr Helen Stanyon	
	Mr Christian Matheou	
	Mr Joe Barritt	

Abstract

The type II secretion system (T2SS) is the major terminal branch of the general secretory pathway in Gram negative bacteria. It is composed of an outer-membrane secretin, periplasmic pilin-like subunits extending from the inner and outer-membranes, an inner-membrane platform, and an associated cytoplasmic ATPase. The secretin is the only essential component in the outer membrane and is formed by a dodecameric torus of protein subunits. Secreted enzymes and virulence factors pass through the gated pore in the centre of the toroidal secretin dodecamer and out into the extra cellular milieu.

In some species of bacteria there are auxiliary proteins that are important for correct assembly and formation of a functional T2SS. One such protein is GspB which, in the plant pathogen *D. dadantii*, aids transport of the secretin from the inner-membrane, through the periplasm and on to the outer-membrane. In the absence of GspB, the secretin miss-locates to the inner membrane and a functional T2SS is not assembled.

GspB is one of four inner-membrane proteins of the T2SS that have a single membrane-spanning helix and periplasmic domain, the others are GspC, GspL and GspM. The hypothesis is therefore that the periplasmic domain of GspB reaches out from the inner-membrane platform into the peptidoglycan-filled periplasm to facilitate the transport of the secretin subunits to the outer-membrane or to stabilize the secretin subunits once they have reached the outer-membrane.

In this Thesis, I report the crystal structure of the periplasmic domain of the auxiliary protein GspB. The structure is similar to the homology region of one of the other four single transmembrane helix proteins from the inner-membrane platform, GspC.

The homology region domain of GspC had been previously shown to interact with the N-terminal domain of the secretin subunit and these observations suggests that GspB may have taken over or may augment the function of GspC. Here, NMR spectroscopy was used to confirm an interaction between GspB and the N-terminal secretin domain and a model of the interaction is proposed on the basis of the available spectroscopic and biochemical measurements.

Contents

List of Figures

List of Tables

Abbreviations

Å	Ångstrom
µl/µM/µS	Micro-liter/micro-molar/micro-second
ATP	Adenosine-5'-triphosphate
B-factor	Temperature factor
CCP4	Collaborative computational project 4
CD	Circular dichroism
COOT	Crystallographic object-oriented toolkit
CV	Column volumes
Da/kDa	Dalton/kilo Dalton
DLS	Dynamic light scattering
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EHEC	Enterohemorrhagic E. coli
ESI	Electrospray ionisation
ETEC	Enterotoxigenic Escherichia coli
F_{cal}	Calculated structure factor amplitude
F_{obs}	Observed structure factor amplitude
FFT	Fast Fourier transform
Gsp	General secretory pathway
gsp	Genes encoding Gsp proteins
GST	Glutathione S-transferase
hkl	Miller indices

HSQC	Heteronuclear single quantum coherence
IM	Inner membrane
IMP	Inner membrane platform
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kd	Dissociation constant
LB	Luria broth
MALDI	Matrix-assisted laser desorption/ionisation
MCS	Multiple cloning site
min.	Minute
ml/mM/mS	milli-liter/milli-molar/milli-second
MS	Mass spectrometry
nl/nM/nS	nano-liter/nano-molar/nano-second
NMR	Nuclear magnetic resonance
NTD	N-terminal domain
OM	Outer membrane
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
PEG	Polyethylene glycol
Pel	Pectate lyase
Pem	Pectin methylesterase
pH	The power of hydrogen (-log10 of the hydrogen ion concentration)

PMSF	Phenylmethethylsulfonyl fluoride
R-factor	Crystallographic residual for working set of reflections
R-free	Crystallographic residual for test set of reflections
R_{merge}	Extent of agreement among multiple measurements of the same reflections
R_{sym}	Measurement of agreement among independent measurements of the symmetry related reflections in a crystallographic data set
RMSD	Root mean square deviation
R_{pim}	Precision indicating merging R-factor
RNA	Ribonucleic acid
S75/200 column	Superdex 75/200 HR 10/30 column (GE Healthcare)
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
Sec	Second
T2SS	Type two (II) secretion system
T3SS	Type three (III) secretion system
T4P/T4PS	Type four (IV pili/type IV pilus assembly system
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TM	Transmembrane
Tris-HCl	2-Amino-2-hydroxymethyl-propane-1,3-diol-hydrochloric acid
UV	Ultraviolet

Chapter 1: Introduction to the Type II Secretion System and the role of GspB

1.1 Bacterial Classification and Architecture of the Cell Envelope

Structurally, bacteria can be divided in to two groups; those possessing an outer membrane (Figure 1.1a) and those without (Figure 1.1b). Their differential staining using Gram's method has been an important way to distinguish and classify bacteria based upon this major structural difference, since its discovery in 1884. Gram negative bacteria possess both an inner and outer membrane between which lies a significantly thinner layer of peptidoglycan and thus they do not retain the crystal violet stain used in the Gram staining method. Gram positive bacteria however, have a much thicker layer of peptidoglycan and hence retain the crystal violet stain much more strongly (hence they stain positively using this method).

Although bacteria are traditionally divided into the two main groups (Gram positive and negative) based on their Gram stain retention, this classification system can be ambiguous as it can refer to three different aspects, that of staining result, cell-envelope organisation, or taxonomic group, which do not necessarily all agree for some bacterial species. In this thesis, I use the term Gram negative to refer to bacteria which possess an inner and outer membrane, between which lies the peptidoglycan containing periplasmic space. The periplasm is a true cell compartment, filled with periplasmic fluid, which has a gel-like consistency (Hobot *et al.* 1984). Proteins located in the periplasm are all specifically targeted with a signal sequence.

These bacteria are also commonly referred to as diderm bacteria (Gupta R., 1998). The diderm bacteria can also be further differentiated between simple diderms lacking lipopolysaccharide, the archetypical diderm bacteria, in which the outer cell membrane contains lipopolysaccharide, and the atypical diderm bacteria, in which outer cell membrane is made up of mycolic acid (Gupta R., 2011).

In Gram negative bacteria, the inner membrane is comparatively symmetrical with both leaflets of the bilayer composed of phospholipid. The outer membrane however, is frequently asymmetrical, possessing mycolic acid or lipopolysaccharide in the outer leaflet and phospholipids in the inner leaflet (Mühlradt *et al.* 1975).

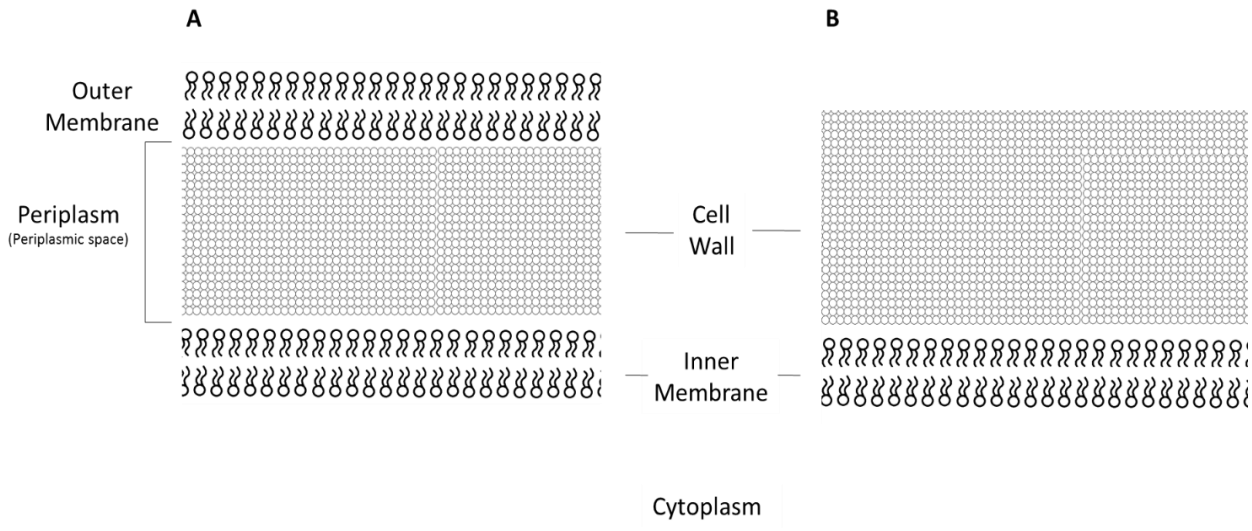


Figure 1.1 Bacterial membranes
The organization of membranes at the cell surface in Gram negative (A) and Gram positive (B) bacteria. Note the substantially thicker cell wall (made of peptidoglycan) found in Gram positive bacteria, which results in them retaining the crystal violet stain much more strongly than Gram negatives.

Bacterial secretion systems allow them to interact with their environment by secreting proteins which may act as toxins or enzymes and can make the external environment more favourable by degrading compounds as a source of nutrients or inducing uptake into host cells. Bacterial secretion system also contribute to efflux pumps, which have a role in the establishment of bacteria in the host and contribute to drug resistance (Piddock, 2006). All bacterial proteins are transcribed in the cytoplasm and must therefore cross various membranes in order to reach the extracellular milieu. The system in Gram negative bacteria is necessarily more complex than that of Gram positives, as they possess 2 lipid bilayers which compounds and proteins must cross in order to enter the external environment. Numerous different systems have evolved to facilitate the passage of proteins and other compounds across bacterial membranes.

1.2 Secretion Systems in Gram negative bacteria

At least nine different secretion systems are known these are called the Type 1 (I) through 9 (IX) secretion systems (Tseng *et al.*, 2009; Desvaux *et al.*, 2009; McBride & Zhu, 2013) of which the first seven are the best characterized to date (Figure 1.2). Of these, Types I, III, IV and VI secrete proteins across both membranes directly in one step. For the other systems (Types II, V and VII), proteins intended for secretion must first cross the inner membrane into the periplasmic space by either the SEC or TAT pathways (Pugsley, 1993) and then be recognised and targeted for secretion across the outer membrane. An N-terminal signal sequence targets proteins to the periplasm.

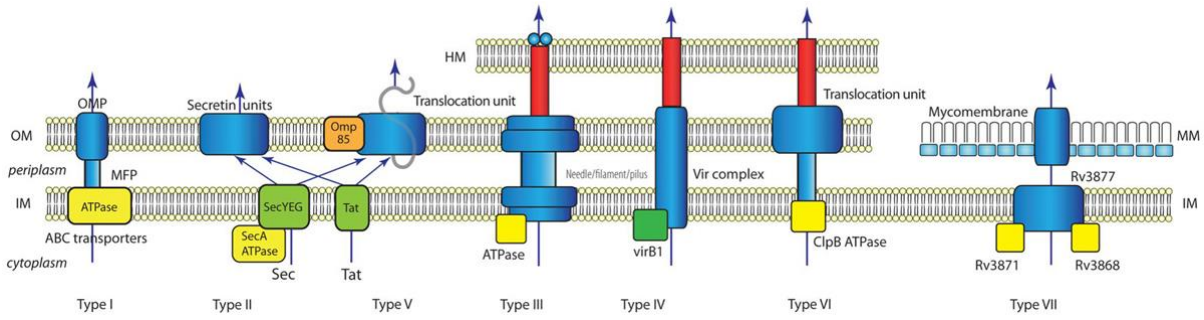


Figure 1.2 Summary of known bacterial secretion systems
Simplified cartoons of the 7 bacterial secretion systems. HM: Host membrane; OM: outer membrane; IM: inner membrane; MM: mycomembrane; OMP: outer membrane protein; MFP: membrane fusion protein. ATPases and chaperones are shown in yellow. [This figure is taken from Tseng *et al*. 2009]

Proteins which cross the inner membrane have a signal sequence, whose function is to help direct the protein to the cytoplasmic membrane. Signal peptides in bacteria fall exclusively at the amino terminal and have a long hydrophobic region (H domain) that is usually preceded by one or more positively charged residues in a short, generally hydrophilic region (the N domain). Shortening or introducing charged or strongly polar residues into the H domain or eliminating the basic amino acids from the N domain usually reduces export efficiency (Pugsley 1993).

Signal peptides are processed and removed during translocation across the cytoplasmic membrane or by signal peptidases present in the periplasm so are therefore presumed unlikely to influence subsequent events in the periplasm or further transport across the outer membrane.

Proteins translocated across the inner membrane by the TAT pathway arrive in the periplasm fully folded, whilst SEC substrates are exported as linear polypeptides to adopt their native 3 dimensional folded state in the periplasmic space. The periplasmic space forms a discrete cellular compartment with its own specific properties and REDOX potential. This is an important feature as proteins translocated across the inner membrane by the SEC pathway will fold in the periplasm aided by specific ions and proteins including chaperones (Pauwels *et al.* 2006), Dsb oxido-reductases (Pugsley 1992) and calcium (Jones *et al.* 2002). Proteins found in the periplasm also have a variety of other functions including nutrient binding; transport, folding, degradation, substrate hydrolysis, peptidoglycan synthesis, electron transport, and xenobiotic metabolism (Klein *et al.* 2005)

1.3 Type II Secretion System and its Role In Disease

The Type II Secretion System (T2SS) is unique in that it allows proteins which have adopted their 3D conformation in the periplasm to be secreted in a fully folded sate. Genes encoding the components of the T2SS machinery were first discovered in the γ-proteobacterium *Klebsiella oxytoca* and were named with the prefix 'Pul' as they were found necessary for the secretion of pullulanase, an alpha-glucanase enzyme (Pugsley and Reyss, 1990). Since its discovery, T2SS genes have been discovered in numerous other genera of γ-proteobacteria including *Acinetobacter, Aeromonas, Erwinia,, Escherichia, Idiomarina, Klebsiella, Legionella, Methylococcis, Photobacterium, Pseudomonas, Shewanella, Vibrio, Xanthamonas, Xylella* and *Yersinia* (Cianciotto 2005).

Subsequently T2SS genes have been discovered in the α-proteobacteria (*Bradyrhizobium, Caulobacter, Gluconacetobacter, Mesorhizobium*), β-proteobacteria (*Azoarcus, Burkholderia, Chromobacterium, Ralstonia*), δ-proteobacteria (*Bdellovibrio bacteriovorus, Geobacter sulfurreducens*) and ε-proteobacteria.

Bacteria that use T2SS and cause disease in humans include both direct (*e.g. V. cholerae*) and opportunistic pathogens (*e.g. K. oxytoca*). Similarly, plant pathogens can have broad host ranges (*e.g. R. solanacearum*) or infect only specific plants (*e.g. E. carotovora*). Pathogenic bacteria with functional T2SS cause a variety of diseases; for example, in humans these range from pneumonia (*L. pneumophila, P. aeruginosa*) to urinary tract infections (*E. coli*) to dysentery (*V. cholerae*). The importance of T2SS to bacterial pathogenesis would appear not to be restricted to a particular site of infection or pathogenic process and it is likely bacteria use the T2SS to boost pathogenesis in various ways even when considering a single species and disease.

Pathogenic processes facilitated by the T2SS include tissue destruction, cytotoxicity, adherence, spread and transmission. It has been shown in *L. pneumophila*, which infects macrophages, that the T2SS can facilitate pathogenesis by promoting bacterial growth in intracellular niches (Rossier *et al* 2004). In other cases of human infection, there are also effectors that subvert the innate immune system; for example, the StcE protease of enterohemorrhagic *E. coli* cleaves components of the complement blood clotting pathway (Grys *et al.* 2005) and the ProA/Msp protease of *L. pneumophila* alters polymorphonuclear leukocyte function (Rossier *et al* 2004).

When present in bacterial pathogens, T2SS systems have usually proven to be relevant and crucial for pathogenesis. Often, the T2SS works along with other secretion systems to achieve full virulence; for example, Type II and IV secretion systems are known to function in *L.*

pneumophila (Rossier *et al* 2004), Type II and III secretion systems function in *X. campestris* (da Silva *et al.* 2002), and Type I, II, III, V and VI secretion systems are present in *P. aeruginosa* (Ma *et al.* 2003; Hauser, 2009; Jyot *et al.* 2011; Hachani *et al.* 2011).

The plant pathogen *Dickey dadantii* contains the genes for functional Type II (Shevchik & Condemine, 2000) and Type III (Yang *et al*. 2008) secretion systems. These different secretion systems make *D. dadantii* a potent pathogen of numerous plant species. The different systems allow the bacteria to modify both its environment in the extra cellular space of plants (by degrading the cell wall as a source of nutrients) and to allow it to invade plant tissues directly (T3SS to directly inject effector proteins to subvert plant immune responses).

1.4 Similarities between the T2 and T4 Secretion Systems

Evidence for the ancestral relationship between the T2SS and the type IV pilus (T4P) system, involved in cell attachment and/or twitching motility, has grown significantly over recent years. It is now clear they share a common progenitor and have significant structural and functional similarities that point to a common origin (Ayers *et al.* 2010). This is similar to the relationship hypothesised between the bacterial flagellum and the type III secretion system. Although the functional details of the T2S and T4P systems are now emerging, there remain significant gaps in our understanding of how they are dynamically assembled in the cell envelope.

The T2SS is also highly similar to the analogous contact independent type IV secretion system (T4SS). The T4SS exchanges macromolecules between cell and external milieu as does the T2SS (Zechner *et al.* 2012). This includes the pertussis toxin liberation system (Ptl), which secretes a complex protein toxin across the cell envelope of *B. pertussis* to the cell exterior (Weiss *et al.* 1993) Similarly to the T2SS, translocation of pre-proteins intended for export, occurs as individual subunits, which cross the IM via the Sec secretion pathway (Farizo *et al*. 2002).

Once they reach the periplasm, the subunits associate with the OM and disulphide bond formation stabilizes the holotoxin. For the assembled holotoxin to traverse the OM, it must enter the Ptl secretion channel from the periplasm. The route of entry is not known, nor is it understood how this process is controlled (Zechner *et al.* 2012).

Among the many challenges faced in the pursuit of a unified model for T2S and T4P function is the limited usage of standardized nomenclature for functionally and structurally equivalent proteins, both within the respective systems and across the T2ss & T4P systems. In part, this problem has arisen due to the identification of T4P- and species-specific components, for which there are no obvious equivalents in the Gsp scheme as originally proposed (Ayers *et al.* 2010).

1.5 Architecture of the T2SS and Essential Components

Nomenclature of the T2SS components can be confusing as there are many species specific names for the individual protein components, as the homologous systems were discovered simultaneously and their relationship to one-another was not fully established at the time of naming. For simplicity the prefix Gsp (general secretory pathway) is used. When referring to a particular protein, the species specific nomenclature is given in superscript. (*E.g.* GspD^{OutD} refers to the *Dickey dadantii* protein OutD, a GspD homologue See Table 1.)

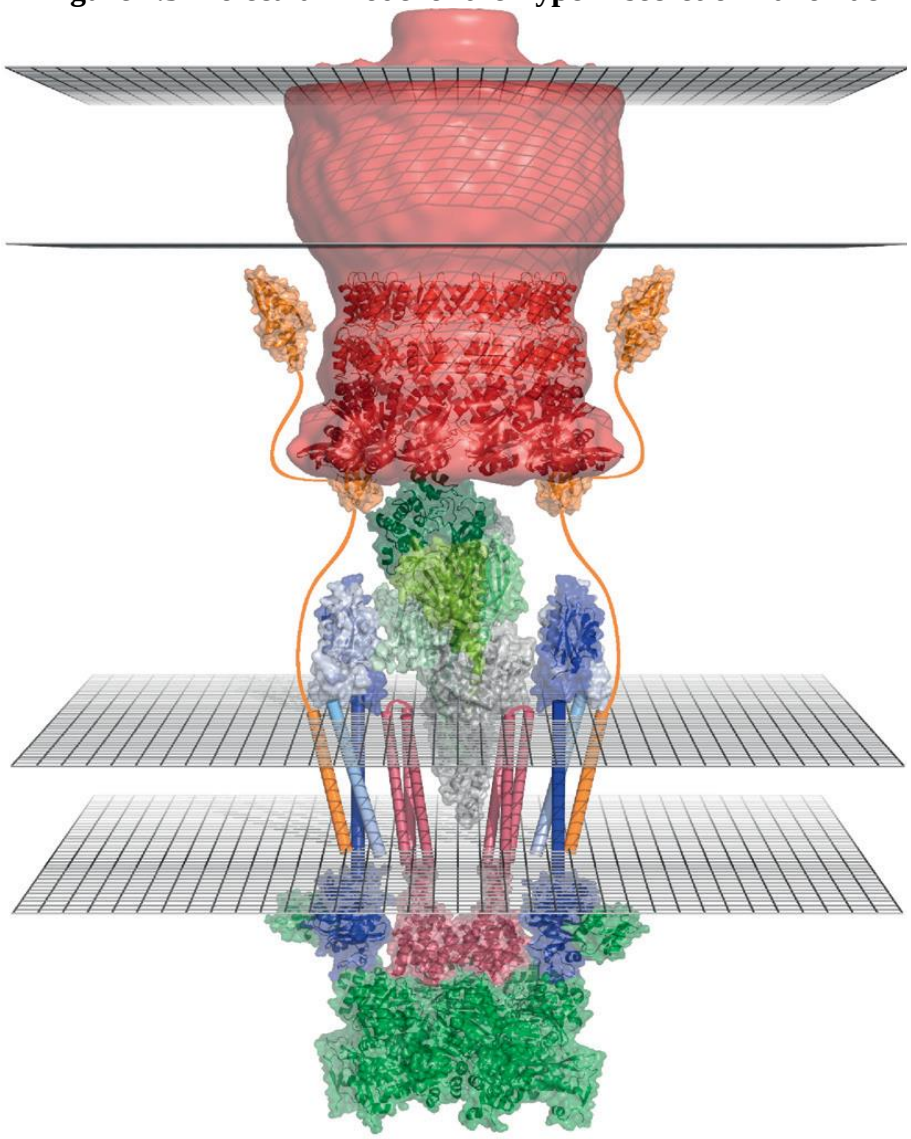
LOCATION	PROTEIN FUNCTION	*Escherichia coli*	*Vibrio cholerae*	*Aeromonas hydophila*	*Dickeya dadantii*	*Klebsiella oxytoca*	*Pseudomonas aeruginosa*	*Xanthomonas camperstris*
Outer Membrane	Secretin	GspD	EpsD	ExeD	OutD	PulD	XcpQ	XpsD
	Pilotin	YghG *			OutS	PulS		
Periplasm	Major Pseudopilin	GspG	EpsG	ExeG	OutG	PulG	XcpT	XpsG
	Minor Pseudopilin	GspI	EpsI	ExeI	OutI	PulI	XcpV	XpsI
	Minor Pseudopilin	GspJ	EpsJ	ExeJ	OutJ	PulJ	XcpW	XpsJ
	Minor Pseudopilin	GspK	EpsK	ExeK	OutK	PulK	XcpX	XpsK
	Minor Pseudopilin	GspH	EpsH	ExeH	OutH	PulH	XcpU	XpsH
Inner Membrane	IM Platform	GspC	EpsC	ExeC	OutC	PulC	XcpP	XpsC
	IM Platform	GspF	EpsF	ExeF	OutF	PulF	XcpS	XpsF
	IM Platform	GspL	EpsL	ExeL	OutL	PulL	XcpY	XpsL
	IM Platform	GspM	EpsM	ExeM	OutM	PulM	XcpZ	XpsM
	Peptidoglycan binding	GspA	EpsA	ExeA				
	Unknown		EpsB	ExeB	OutB	PulB		
Cytosol	Secretion ATPase	GspE	EpsE	ExeE	OutE	PulE	XcpR	XpsE

Table 1.1 The location and nomenclature of the major components of the T2SS
The location within the 2 membranes, and the species specific nomenclature of the major components of the T2SS present in *E. coli*, *V. cholera*, *A. hydrophilia*, *D. dadantii*, *K. oxytoca*, *P. aeruginosa* and *X. campestris*.

Type II protein secretion is also known as the main terminal branch of the general secretory pathway (Gsp) or secretin-dependent pathway. The system spans both the inner and outer membrane of Gram negative bacteria and facilitates the final stage of a two-step process in which proteins are first translocated across the inner membrane by the SEC or TAT pathways and then transported from the periplasm to the exterior through an outer membrane secretin pore.

Dependent on the bacteria, the T2SS is composed of 12 to 16 components; GspA to GspO and GspS. This trans-envelope secretion machinery spans the two bacterial membranes and can be roughly divided into three main blocks an inner membrane platform, trans-periplasmic pseudopilus, and outer membrane pore (Korotkov *et al*., 2012), as shown in Figure 1.3.

Figure 1.3 Molecular model of the Type II secretion nanomachine



The model of the **GspD N-terminal domains** (PDB: 3EZJ) is positioned within the *V. cholerae* **GspD EM reconstruction** (red surface) (PDB: 3EZJ and EMD: 1763).

GspC HR domain was aligned to GspD twelve-fold ring using the GspC–GspD co-crystal structure (PDB: 3OSS). GspC PDZ domain (2I4S) is positioned near the GspD N3 domain, just under the OM.

Minor pseudopilins (GspK, dark green; GspHIJ, light greens) were manually added by grafting helices onto *V. cholera* EpsH (PDB: 2QV8) and ETEC GspIJK (PDB: 3CI0) structures

GspF dimer was created with two EpsF domain-1 homodimers (PDB: 3C1Q).

One each of **GspL** and of **GspM** periplasmic domains (PDB: 2W7V; 1UV7) were aligned to the PilO homodimer (2RJZ) to create the GspL–GspM heterodimer (surface from PilO dimer.)

GspE hexamer was modeled by aligning *V. cholerae* EpsE domains (1P9R) onto the homologous domains of the P. aeruginosa PilT hexamer (3JVV).

Stoichiometry and relative arrangement of inner membrane proteins is currently unknown, although a 6:6:6:6:12 assembly of GspE: L: M: C: D is consistent with available data. For model viewing, one GspF dimer, two GspL-M heterodimers, and two GspC chains are shown.

[Taken from McLaughlin *et al.* 2012]

1.6 Inner Membrane

GspC, GspL and GspM all have a single transmembrane helix with GspF the exception having at least three transmembrane helices. GspE is cytoplasmic but associates with the inner membrane components.

GspE

The cytoplasmic ATPase GspE is associated with the inner membrane platform (Camberg, 2007) and is associated with the inner membrane through the cytoplasmic domain of GspL binds to the extended N-domain of the assembly ATPase GspE in a species-specific fashion (Sandkvist *et al* 1995, Abendroth *et al.* 2005) while its periplasmic domain interacts with GspM.

GspE energizes formation of a short pilus, restricted to the periplasm and composed of five pseudopilins GspG, H, I, J and K (Nivaskumar *et al.*, 2014). The dynamic assembly and/or rotation of this pilus may facilitate the passage of folded proteins through the outer membrane pore composed of the secretin GspD.

GspF

The essential component GspF is fully embedded in the IM via the three TM segments (Robert *et al.* 2002, Thomas *et al.*1997). The cytoplasmic domain of GspF probably shares the same 6-helix bundle structure, which has been revealed by X-ray analysis of the N-terminal domain Cyto1 (Abendroth 2009). Yeast two-hybrid and pull-down studies suggest that GspF Cyto1 interacts with GspE and GspL, forming the so called inner membrane platform that also includes GspM. GspF requires GspL and GspE for full stability against proteolytic degradation. Although GspF is crucial in the assembly of a functional T2SS, very little is known about its oligomeric arrangement and structure within the membrane.

GspL

GspL is another essential component of the inner membrane platform and may provide the physical link between the ATPase GspE and the membrane and also regulate its activity as well as allow coupling to pilus assembly. Studies in *X. campestris* show that only the ATP-bound form of $GspE^{XpsE}$ interacts with $GspL^{XpsL}$ (Shiue *et al.* 2006). The last 11 residues of the cytoplasmic domain of GspL are crucial for cardiolipin binding, which stimulates the ATPase activity of GspE (Camberg *et al.* 2007). Notably, $GspL^{EpsL}$ could be cross-linked to the major pseudopilin $GspG^{EpsG}$ in *Vibrio cholerae*, providing further evidence for the role of ATPase in pseudopilus assembly on a molecular basis (Gray *et al.* 2011).

GspM

GspM is a bitopic IM protein that interacts with GspL (Johnson *et al.* 2007). It has two periplasmic domains which both have a ferredoxin-like fold and stabilise each other, forming heterodimers and help ensure the correct cellular localisation (Lybarger *et al.* 2009). Whilst the other bitopic IM protein, GspL, is essential for both fiber assembly and secretion (Campos *et al.* 2013), the deficiency of GspM still allows for assembly of the pseudopilus, which suggests that perhaps overproduction of other components partially compensates for its absence in at least one species, *Pseudomonas aeruginosa* (Durand *et al.* 2005).

Recent studies suggest that GspL and GpsM interact through their TM segments, along with GspC and form dynamic homo- and heterodimers (Lallemand *et al.* 2013). Structural studies reveal that the GspL and GspM homologues of the Type 4 Secretion System (T4SS) of *T. thermophilus* (PilN and PilO respectively) bind the major pilins during pseudopilus elongation (Karuppiah *et al.* 2013), indicating the essential role these IMP proteins play.

GspC

GspC is another bitopic, transmembrane protein anchored to the inner membrane. It has two periplasmic domains, the homology region (HR) near the membrane and the C-terminal PDZ domain that is involved in substrate specificity in the *D. dadantii* T2SS (Bouley *et al.* 2001). The structure of the HR domain, revealed high similarity with the T4SS lipoprotein PilP (Korotkov *et al*., 2011, Gu *et al*., 2012b). The transmembrane segment of $GspC^{OutC}$ mediates essential dimerization in *D. dadantii* (Login and Shevchik 2006), whilst the HR domain interacts with N0 domain of GspD as shown by the recent structural and NMR studies (Gu *et al.* 2011). This contact, between GspC and GspD is essential for function and provides a contact across the periplasm between the platform of the inner membrane and the outer membrane pore. The strength of the interaction between GspC and GspD varies between species and is probably the basis for the species-specificity identified in systematic exchange experiments of Out T2SS components of *D. dadantii* and *P. carotovorum* (Lindeberg *et al.* 1996).

Studies in *D. dadantii* suggest however that this link is flexible and involves multiple dynamic interfaces, and also that the stoichiometry of the C–D complex might vary (Login *et al.* 2010). GspC and the periplasmic flexible domains of GspD are implicated in substrate binding and selection. This has been proposed based on the domain exchange studies in the Out T2SS of *D. dadantii* and *P. carotovorum* (Bouley *et al.* 2001). Further support for the roles of both GspC and GspD in substrate binding comes from in vitro studies showing that $GspG^{XcpP}$ binds specifically to the Xcp system substrate LasB (Douzi *et al.* 2011). The C-terminal PDZ/coiled-coil motif of GspC has been implicated in these protein–protein interactions (Korotkov *et al.* 2006). However, this motif is required for secretion of only a subset of substrates in the *D.*

dadantii Out T2SS (Bouley *et al.* 2001). *In vitro*, cholera toxin binds to the periplasmic secretin vestibule in the absence of GspC (Reichow *et al.*2011).

1.7 Outer Membrane Secretins

GspD

The secretins are a family of proteins which form pore like structures. They are present in the bacterial Type II, III and IV secretion systems (Figure 1.4a) and form megadalton-sized, dodecameric, pore-like structures (Figure 1.4b) in the outer membrane; through which secreted substrates are translocated (Reichow et al., 2010).

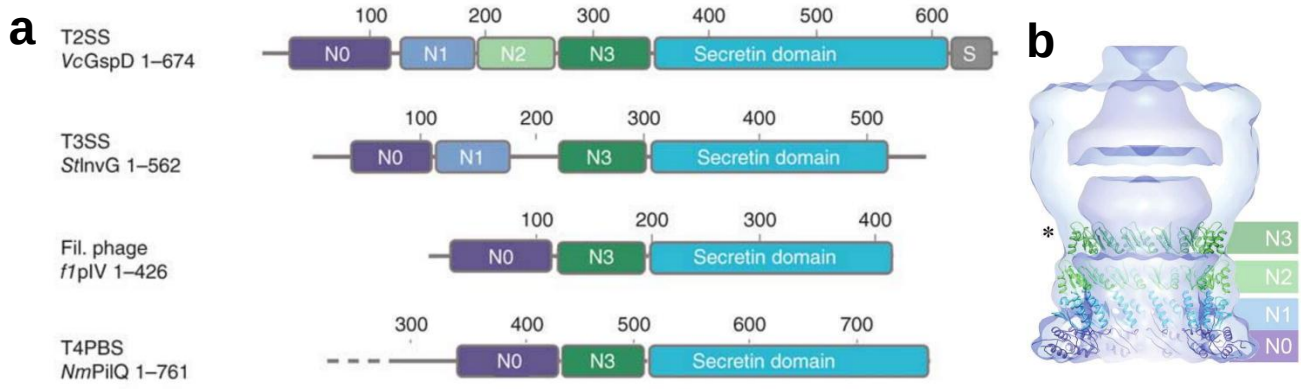


Figure 1.4 Domain architecture and structure of secretins
(a) Schematic representation of the domains within secretins from various secretion and related systems. (**b**) Cartoon showing the 3D structure of a secretin pore and the arrangement of the N-terminal domains within it. [Adapted from Hol *et al.* 2012]

The N-termini of secretins possess serval domains, depending on species, which protrude into the periplasm. In the T2SS, the majority of GspD homologues have 4 N-terminal domains (N0, N1, N2, and N3). The crystal structure of the first three N-terminal domains (N0, N1, N2) of the T2SS secretin GspD from enterotoxigenic *Escherichia coli* has been solved with the assistance of an antibody fragment (PDB: 3EZJ). The structure shows that the three periplasmic GspD domains are arranged into two lobes: a compact N-terminal lobe containing the N0 and N1 domains, and a second lobe containing the N2 domain. The GspD domains N1

and N2 share the same fold, which is different from that of the N0 domain, which interacts with the inner membrane component GspC (Korotkov *et al.*, 2011; Gu *et al.*, 2012a).

The N0 domain reveals a structure similar to the TonB-dependent receptor. The N1 and N2 domains have a fold similar to the eukaryotic type I KH domain, which is typically involved in binding RNA and DNA. However, the positively charged RNA/DNA binding residues are absent in the secretin N1 and N2 domains.

The N3 domain structure is not known, but it has a high degree of sequence similarity with the N1 and N2 domains, which suggests that the N3 domain is likely to have a fold similar to N1 and N2. It is interesting to note that the crystal structure of the N0 and N1 domain of T3SS secretin reveals a different orientation to that seen in the T2SS (Korotkov, Gonen *et al*. 2011). The currently known GspD N-terminal domains structures were both solved with the assistance of modified antibodies, which binds both N0 and N1 and could have a sizeable influence on the domains orientations within the crystal lattice.

To reach the outer membrane and to be correctly assembled, the large secretin protomers (~80 kDa) have to pass through the densely packed periplasm and the peptidoglycan mesh network. A special class of carrier proteins, called pilotins, ensures the transit of secretins through the periplasm and their assembly into the outer membrane.

In the T2SS the secretin, GspD, is an essential, non-interchangeable component and is the only integral outer membrane component (Hardie *et al.* 1996). It exists as a dodecameric oligomer with each subunit composed of a C-terminal transmembrane secretin domain, a pilotin interacting S-domain and three N-terminal domains termed N0 through N3.

The C-terminal secretin domain is the most conserved domain across all secretins and is responsible for oligomerisation (Guilvout *et al*. 2011). At the time of writing the only structural

information comes from cryo-electron microscopy and no high resolution (X-ray / NMR) structures are available, but it is presumed to consist of a transmembrane β-barrel type structure.

The N-terminal domains are distinct, and less conserved across different species. These domains extend into the periplasm and are involved in a series of interaction with both secreted proteins and proteins of the inner membrane platform. The N0-N1 domains have been crystalised in isolation (PDB: 3EZJ) and in complex with the HR domain (PDB: 3OSS). The N0 domain alone has also been crystalised, with P6 symmetry, and its structure modelled to form the C12 symmetry necessary to form a dodecamer.

These structures revealed the N0 fold to be similar to the TonB-dependent receptor. The N1 and N2 domains are similar to the eukaryotic type 1KH domain which is typically involved in DNA/RNA binding, but these residues are absent in N1 and N2 and are thus far not implicated in such a role.

The S-domain of GspD is responsible for the direct interaction with the pilotin GspS (Shevchik and Condemine 1998). The pilotin acts as a chaperone, stabilising GspD and facilitating its transport across the periplasm from the inner membrane to the outer membrane in partnership with the Lol pathway (Collin Guilvout 2011).

1.8 Assembly and biogenesis of the T2SS

In one sense, assembly of the T2SS has been thoroughly (although largely indirectly) studied, given that much is known about how proteins are translocated into or across the inner membrane, and that many of the proteins of the T2SS are integral to the inner membrane (Facey & Kuhn, 2010; Froebel *et al.*, 2011). After being assembled into this membrane via the Tat and

Sec/SRP systems, protein–protein interactions dictated by structure might then be likely to result in the spontaneous self-assembly of the inner membrane platform of the system (Dalbey *et al.,* 2011). In this sense much "assembly" of the T2SS would result from the formation of complexes driven by the inherent shape and biophysical properties such as hydrophobicity and electrostatic character of the proteins involved. Beyond the critical involvement of the inner membrane platform, these processes may or may not require additional factors beyond those comprising the complexes themselves.

However, understanding assembly of the T2SS is complicated by the fact that the system resides in both membranes and spans the entire envelope of the cell. This means that components belong to the inner and outer membranes, and some interact with both membranes. This trans-envelope organisation requires additional assembly processes. For example, the movement to and assembly of the secretin in the outer membrane, the establishment of contact between the inner membrane components and the secretin, the growth of the pseudopilus from the inner membrane to the secretin in the outer membrane, and the integration of the entire apparatus with the peptidoglycan net that encircles the cell and resides within the periplasm.

1.9 Auxiliary Components

Pilotins

In many cases, the assembly of secretins in the outer membrane is partly directed in some species by pilotin proteins (Rehman S., *et al.* 2013). These small outer-membrane lipoproteins, assist their cognate secretins in targeting to the outer membrane or in formation / stabilisation of the secretin multimers (Hardi K. R., *et al.* 1999). The C-terminal region in the T2SS and T3SS secretins interacts with their corresponding T2SS and T3SS pilotins, shown as the S-domain in Figure 4a.

Interaction between the secretin GspD and its lipoprotein pilotin is essential of translocation of the secretin to the outer membrane in serval species including *Klebsiella oxytoca* and *Dickeya dadantii.* The pilotins function by directly binding to the unstructured C-terminus of cognate secretins (Rehman et al., 2013). Two structurally dissimilar (unrelated) families of pilotins have been identified in various T2SSs. The structure of the paradigm pilotins from OutS/PulS family, also called GspSα, comprise an arrangement of four α-helices, while recently identified pilotin AspS of GspSβ family is folded in an α/β domain (Gu et al., 2012b; Dunstan et al., 2013).

GspA

In addition, *Erwinia, Aeromonas, Vibrio* and some other genera possess two other auxiliary T2SS components, GspA and GspB, necessary for the correct assembly of their appropriate secretin multimers in the outer membrane (Strozen *et al.*, 2011). GspA possesses a cytoplasmic ATPase domain followed by a transmembrane segment and a large periplasmic region that contains domain A1 of unknown function and a peptidoglycan-binding domain (Li & Howard, 2010). The peptidoglycan-binding domain of GspA belongs to the Pfam family PF01471 and is composed of three α-helices (Martynowski *et al*., 2013). In *Aeromonas* and *Vibrio*, GspA and GspB interact together and form an inner-membrane complex that binds peptidoglycan and is necessary for the assembly of the secretin (Li & Howard, 2010). In some other bacteria, such as *Dickeya*, *Klebsiella* and *Pectobacterium*, only the GspB component is present whilst GspA is apparently lacking. In this case, the *gspB* gene neighbours that of the pilotin gspS, but the two genes are transcribed independently.

GspB

Regardless of the presence or absence of a cognate GspA component, all the GspB proteins share a common composition. GspB is comprised of a short cytoplasmic region, a single

transmembrane segment and a large periplasmic region (Condemine & Shevchik, 2000). OutB is a component of the *Dickeya dadantii* (previously *Erwinia chrysanthemi*) Out secretion machinery. Homologues of OutB have been described in two other bacteria, *Klebsiella oxytoca* and *Aeromonas hydrophila*, but their requirement in the secretion process seems to be different. Study of OutB topology with the BlaM topology probe suggests that it is an inner-membrane protein with a large periplasmic domain. However, fractionation experiments indicate that it could be associated with the outer membrane through its C-terminal part.

The secretion deficiency of an *D. dadantii* outB mutant can be reversed by the addition of an inducer of the kdgR regulon. It was shown that this effect results from the increased expression of the secretin OutD and that secretion can be restored in an outB mutant by introducing the outD gene on a plasmid and overexpressing OutD protein. Several experiments suggest an interaction between OutB and OutD. In *D. dadantii*, the presence of OutD stabilizes OutB. OutD expressed in *Escherichia coli* can be protected from proteolytic degradation by the co-expression of OutB. This effect does not require the N-terminal, transmembrane segment of outB. OutB can be cross-linked with OutD by formaldehyde. These results indicate that OutB could act with OutD in the functioning of the Out secretion machinery.

A recent study in *Aeromonas hydrophila* suggested that GspB interacts with the periplasmic region of the secretin (Vanderlinde et al. 2014). However, the precise function of GspB remains to be elucidated. Furthermore, no structural data are available for GspB and there has been no clear structural homologue from sequence searches.

1.10 Aims and objectives of this study

The aims of this study are to determine the molecular structure of the inner membrane protein GspB homologue OutB from *Dickeya dadantii* and determine the atomic details of its possible interaction with the outer membrane secretin OutD. This will hopefuly illuminate important

details in the assembly of the secretin multimers and the Type 2 Secretion System in general, providing insights in to how this system can be targeted in bacteria which possess homologues systems and are pathogenic to human.

In this thesis, it was possible to address these aims by exploring crystallization of the periplasmic domain of OutB from a number of constructs and solving the crystal structure. This work is described in Chapter 3. The interaction between OutB and the secretin OutD was explored using cross-linking studies and NMR, which are discussed in Chapter 4 and 5, respectively.

Chapter 2: Methodological Introduction

2.1 Molecular biology

2.1.1 Cloning of OutB

DNA encoding the C-terminal region (the periplasmic domain) of *D. dadantii* GspB, residues 112 to 220, was amplified from the then *Erwinia chrysanthemi* strain 3557 using suitable primers and then subcloned into the pGEM-T Easy (Promega) vector. This allowed for easy amplification of the cloned gene so that it could be sequenced and exact sequence confirmed. The insert was then subsequently cloned into the expression vector pGEX-6P-3 (G E Healthcare; Appendix A1). This plasmid encodes a Glutathione S-transferase (GST) affinity tag at the N-terminal which is cleavable at the amino acid sequence LEVLFN|GP by the patented PreScission™ protease (GE Healthcare).

2.1.2 Cloning OutD

Sequence alignment of $GspD^{OutD}$ against GspD from different bacteria shows that the N-terminal portion of its periplasmic region contains the 3 N-terminal domains possessed by many secretins termed N0, N1, N2, and N3. DNA encoding the highly conserved region of the periplasmic N0 domain of *D. dadantii* $GspD^{OutD}$-N0, residues 28 to 112, was subcloned into the pGEM-T Easy (Promega) vector, its sequence confirmed, and then the insert cloned into the pET-14b vector (Appendix A2) encoding a cleavable C-terminal hexahistidine tag for purification with nickel ion affinity chromatography. This vector is also under the control of the *LacI* operon and inducible by the addition of IPTG.

2.1.3 pET14b Vector (Novagen)

The pET expression system (Novagen) is a system developed for the cloning and expression of recombinant proteins in *E. coli*. Target genes are cloned into pET plasmids under control of strong bacteriophage T7 transcription. Expression is induced by providing a source of T7 RNA polymerase in the host cell. T7 RNA polymerase is so selective and active that, when fully induced, almost all of the cell's resources are converted to target gene expression (Novagen pET System Manual). The desired product can comprise more than 50% of the total cell protein a few hours after induction.

It is also possible to attenuate expression levels simply by lowering the concentration of inducer. Decreasing the expression level may enhance the soluble yield of some target proteins. Another important benefit of this system is its ability to maintain target genes transcriptionally silent in the un-induced state. Target genes are initially cloned using hosts that do not contain the T7 RNA polymerase gene, thus eliminating plasmid instability due to the production of proteins potentially toxic to the host cell.

Once established in a non-expression host, target protein expression is initiated by expression by the host of a chromosomal copy of the T7 RNA polymerase gene under lacUV5 control; which expression is induced by the addition of IPTG to the bacterial culture. The pET14b plasmid specifically encodes an N-terminal hexahistidine tag, with a thrombin cleavage site between the His-tag and the target protein of interest.

2.1.4 pGEX-6P-3 Vector (G. E. Healthcare)

The pGEX vectors are designed for inducible, high-level intracellular expression of genes or gene fragments as fusions with *Schistosoma japonicum* GST. Expression in *E. coli* yields

tagged proteins with the GST moiety at the amino terminus and the protein of interest at the carboxyl terminus.

GST occurs naturally as a 26 kDa protein that can be expressed in *E. coli* with full enzymatic activity. Tagged proteins that possess the complete amino acid sequence of GST also demonstrate GST enzymatic activity and can undergo dimerization similar to that observed in nature. The crystal structure of recombinant *S. japonicum* GST from pGEX vectors has been determined and matches that of the native protein.

Purification of GST-tagged proteins is based on the affinity of GST to the glutathione ligand imobilised to a matrix. The binding of a GST-tagged protein to glutathione is reversible, and the protein can be eluted under mild conditions by the addition of reduced glutathione to the elution buffer. This provides a mild purification process that does not affect a protein's native structure and function.

Cleavage of the protein from the GST tag can be achieved using a site-specific protease whose recognition sequence is located immediately upstream from the multiple cloning site (MCS) of pGEX plasmids. Tagged proteins can be detected using colorimetric or immunological methods. PreScission Protease is a fusion protein of glutathione S-transferase (GST) and human rhinovirus (HRV) type 14 3C protease (Walker *et al*. 1994). The protease specifically recognizes a subset of sequences which include the core amino acid sequence Leu-Phe-Gln/Gly-Pro cleaving between the Gln and Gly residues. Substrate recognition and cleavage are likely to be dependent on primary structural signals as well as the secondary and tertiary structures of the fusion protein. Since the protease is fused to GST, it is easily removed from cleavage reactions using Glutathione Sepharose™ 4B.

Fusion proteins produced from pGEX-6P-1, pGEX-6P-2 and pGEX-6P-3 are cleaved by PreScission Protease between the GST moiety and the cloned fusion partner. The molecular weight of PreScission Protease is approximately 46 kDa.

2.1.5 Mutagenesis

Mutagenesis was performed to:

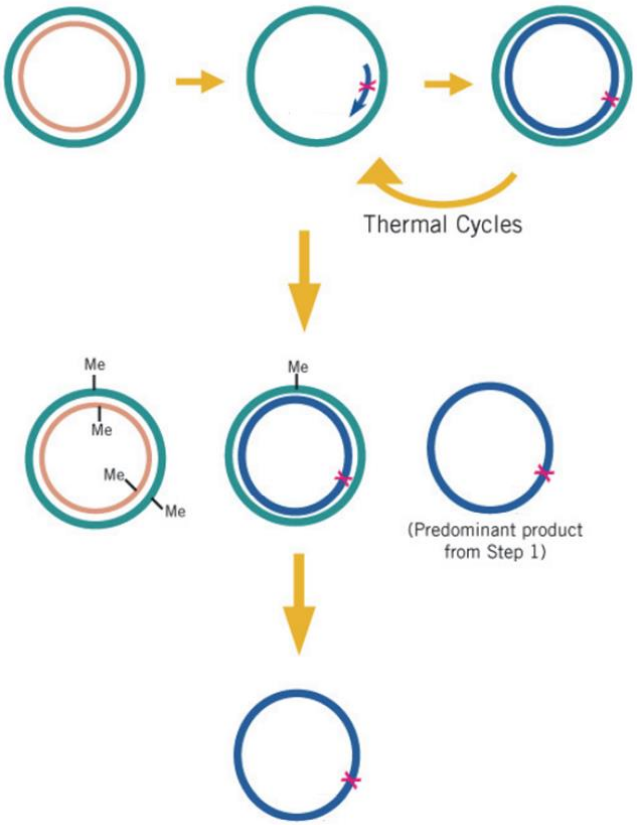
1. Introduce premature stop codons and from C-terminally truncated mutants.
2. Perform point mutations to introduce cysteine residues at positions where proteins were believed to be interacting and thereby trap the complex using disulphide bond formation.

In both cases, this was achieved using the Quick Change Lightening II Mutagenesis kit (Agilent) with primers designed accordingly (See Appendix B). Mutant DNA was transformed in to XL10 ultra competent cells (Agilent) and plasmid DNA extracted by MiniPrep (Qiagen). This was then sequenced (MWG Eurofins) to check incorporation of the stop codon generated via point mutation.

The rapid, simple, three-step method is outlined in Figure 2.1. A single mutagenic oligonucleotide is required to generate a single mutagenic site, using a double-stranded, non-mutated DNA template. A complimentary oligonucleotide incorporating the point mutation and with a GC content >45% must first be synthesized. Temperature cycling then allows for melting of the double stranded template, cooling to allow annealing of the mutagenic primer, and then an extension/synthesis period whereby the mutagenic primer is incorporated into a new plasmid.

DpnI, a type II M restriction endonuclease, is then used to recognize and cut exclusively the template, methylated DNA. Resulting in only newly synthesized, mutagenic plasmids. This

process allows for mutagenesis to be performed in approximately two hours, plus an overnight transformation of the newly generated DNA construct.



1. Thermal cycling

- Denatures DNA template
- Allows mutagenic primer to anneal
- Extend primers and ligate nicks with *pfu* fusion enzyme blend.

2. DpnI digestion of the template

- DpnI exclusive digests methylated and hemimethylated DNA

3. Transform mutated ssDNA into

- XL10 Gold ultracompetent cells.

Figure 2.1 Cartoon and description of the Quick Change Lightning II™ protocol to generate mutations in plasmid DNA

2.1.6 Production and Purification of target proteins

Expression constructs were transformed in to *E. coli* strain BL21(DE3) (Bioline), for protein expression. This strain of competent cells was used in expression of all protein constructs and contains the T7 polymerase gene under the control of the LacUV5 promoter. Addition of IPTG (isopropyl-β-D-thiogalactopyranoside) induces constitutive expression of T7 polymerase. For Expression vectors under the control of a T7 promoter (pGEX and pET used in this work), the polymerase transcribes the T7 polymerase mRNA at a high copy number resulting in over-expression of the desired protein.

2.2 Biochemistry

2.2.1 SDS PAGE Gel

Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE), allows proteins to be separated and visualised according to their electrophoretic mobility, a function of the length of a polypeptide chain (proportional to its mass) and its charge. This can be used in a qualitative way to assess the expression and purification of proteins.

According to the size range of the proteins to be resolved, 12% and 18% polyacrylamide gels were used. These were prepared according to standard molecular biology protocols. Self-cast gels were made, adding TEMED just before pouring the gel mixture into the Bio-Rad gel cast (Bio-Rad). The resolving gel was poured into the cast first, followed by the stacking gel. A comb for 10 or 15 wells was inserted, before the gel was left to set.

2.2.2 UV/Vis

Protein concentration was determined directly by spectroscopic analysis using UV-Vis spectrophotometer (HITACHI U-3010). Protein absorbance at 280nm was measured.

Concentration was determined according to:

Equation 2.1 $A_{280} = \varepsilon \cdot c \cdot l$

Where *c* is the molar concentration, A280 is the absorbance at 280nm, ε is the molar extinction coefficient, typically calculated using the ExPASy ProtParam webserver (http://web.expasy.org/protparam/) and *l* is the path-length of the cuvette.

The absorbance ratio of A260/A280 was measured to check DNA contaminations. An A260/A280 ratio below 0.7 was considered to have no DNA contamination.

2.2.3 Circular Dichroism

Circular Dichroism (CD) is an excellent method to give an indication of the secondary structure of the protein. Although it is not possible to give detailed residue-specific information as obtainable from NMR and crystallography, it can provide secondary structure information using only small amounts of protein and it is extremely sensitive to changes in secondary structure and can also be used to monitor the conformational changes, for instance as a function of temperature, pH, or ionic strength (Martin *et al* 2008).

CD measures the difference in the absorption of right (A_R) and left circular polarised light (A_L). Therefore, for a sample to be CD active it must be chiral. Amino acids are optically active as they contain α-carbons with four different groups attached. The chiral character of a sample allows for the differential absorption of the L- and R- circularly polarised light (ΔA), this will produce a CD spectrum. The difference in absorption can be related to the molar extinction coefficient using the Beer-Lambert's law (Equation 2.2)

Equation 2.2 $$\Delta\varepsilon = \varepsilon_L - \varepsilon_R = \frac{(A_L - A_R)}{c.l} = \frac{\Delta A}{c.l}$$

Where Δε is the difference in molar extinction coefficient, Δ*A* is the difference in absorbance (A_L - A_R), *c* is the molar concentration and *l* is the path-length of the sample in centimetres.

The relationship between the direct CD measurements (in millidegrees), θ, and Δε, $M^{-1}cm^{-1}$, is highlighted in equation 2.3 below. Please note that the relationship between CD measurement and change in absorbance is θ = Δ*A* . 33,000.

Equation 2.3 $$\Delta\varepsilon = \frac{\theta}{33{,}000.c.l}$$

Where θ is the direct CD measurement in millidegrees, *c* is the molar concentration and *l* is the path-length of the cuvette in centimetres.

The far-UV region spans wavelengths 170 to 250 nm, this region is essential for measuring secondary structural changes of a protein. Far-UV-CD uses the two electronic absorptions of the backbone amide group, the electronic dipole from the π – π* transition at 190 nm and the magnetic dipole from the weaker *n* - π* transition at 210 nm. These transitions dominate the CD spectrum to produce a trace characteristic of the ψ (psi) and φ (phi) torsion angles, therefore, are characteristic of secondary structure of the polypeptide chain. Different structural conformation produce characteristic CD spectrum, as shown in Figure 2.2. Due to the versatile nature of far-UV-CD it has been as a valuable technique for thermal denaturation studies.

Visible Region. This region of the CD spectrum is useful for monitoring metal – ligand binding and it makes use of the wavelengths from 300 to 800 nm. The visible region monitors the difference in the charge transfer, metal, *d* – *d*, and ligand, p – p*, transitions. As with far-UV-CD, a trace is only observable if a metal is bound to a ligand in a chiral environment. Therefore, free metal ions, even with *d* – *d* transitions, are CD silent.

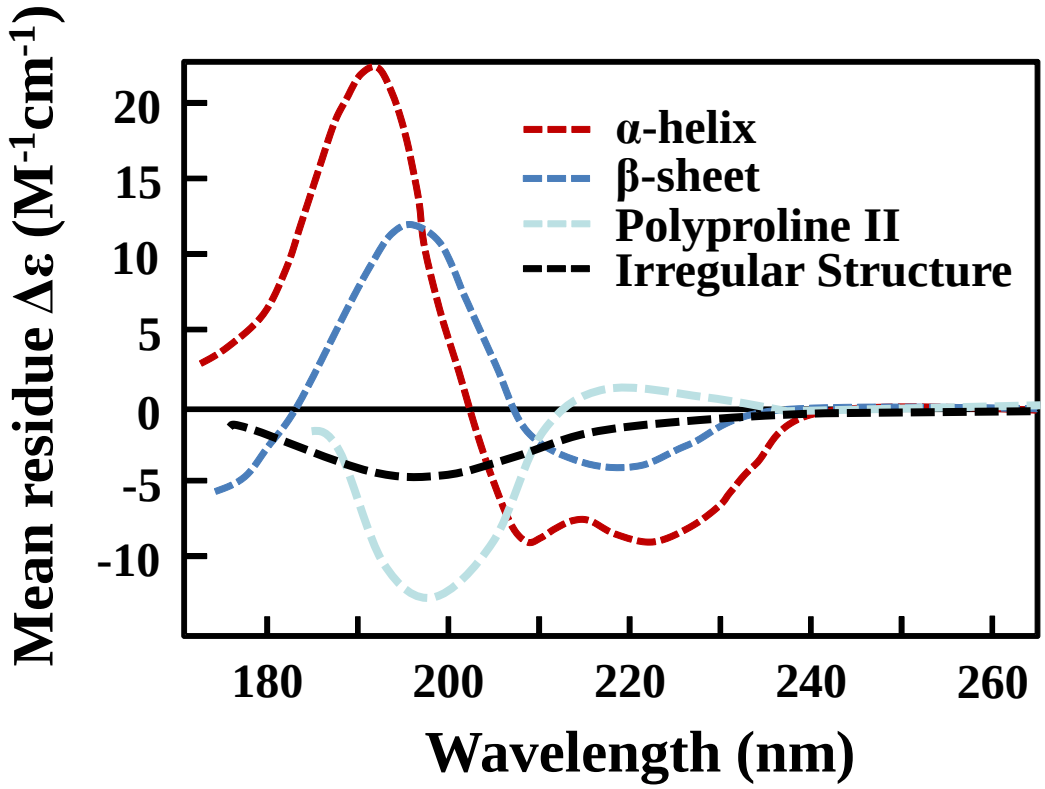


Figure 2.2 Typical CD Traces for secondary structure elements

This far-UV-CD illustration demonstrates the different structural conformations (such as α-helix, β-sheet, polyproline II and irregular structure) and the characteristic traces that are produced. This figure has been adapted from S. Brahms *et al.* (Brahms, S., *et al.* 1980).

2.2.4 Dynamic Light Scattering

Dynamic Light Scattering (DLS), also known as Photon Correlation Spectroscopy (PCS) and Quasi-elastic Light Scattering (QELS), is a technique used to determine the size distribution profile of small particles in suspension. In DLS, a laser beam is directed into a solution and, the light energy is scattered when it meets a particle. The scattered light is detected on a photon detector connected to a correlator.

The scattered light is a function of the hydrodynamic size and the concentration of the sample. Higher concentrations of molecules scatter more light than lower concentrations. Similarly larger molecules scatter more light than smaller molecules at the same concentration.

Time-dependent fluctuations in the scattered light occur because the particles are moving under Brownian motion; these fluctuations are directly related to the rate of diffusion of the molecules in the solution. The correlator compares the intensity of the scattered light at time *t* with that at time t+δt, t+2δt, t+3δt, etc, and thus generates a correlation function. Since small molecules diffuse faster, they lead to a faster decay in the measured correlation function. The measured decay rate is thus related to the hydrodynamic size of the scattering object.

In dynamic light scattering, the speed at which the particles are diffusing due to Brownian motion is measured. This is done by measuring the rate at which the intensity of the scattered light fluctuates when detected using a suitable optical arrangement. The rate of the scattering intensity fluctuation occurring depends on the size of particles. The small particles cause the intensity to fluctuate more rapidly than the large ones. The size of a particle is calculated from diffusion coefficient using the Stokes-Einstein equation (Equation 2.4).

Equation 2.4 $$\mathbf{D} = \frac{k_B T}{6\pi\eta R}$$

2.2.5 Mass Spectrometry

The most relevant form of mass spectrometry as applied to macromolecules, particularly proteins, is electrospray ionization (ESI). This technique, first reported by Masamichi Yamashita and John Fenn in 1984 (Yamashita & Fenn, 1984) is used in mass spectrometry to produce ions using an electrospray in which a high voltage is applied to a liquid to create an aerosol.

It is especially useful in producing ions from macromolecules as it overcomes the tendency of these molecules to fragment when ionized. ESI is different from other atmospheric pressure ionization processes (e.g. MALDI) since it may produce multiply charged ions, effectively extending the mass range of the analyser to accommodate the kDa-MDa orders of magnitude observed in proteins and their associated polypeptide fragments (Ho *et al.* 2003).

Mass spectrometry using ESI is called electrospray ionization mass spectrometry (ESI-MS) or electrospray mass spectrometry (ES-MS). ESI is a 'soft ionization' technique as very little fragmentation occurs. This is advantageous as a molecular parent ion (or more accurately a pseudo molecular ion) is always observed, however very little structural information can be gained from the simple mass spectrum obtained. This disadvantage can be overcome by coupling ESI with tandem mass spectrometry (ESI-MS/MS).

Electrospray ionization is the preferred technique to couple liquid chromatography with mass spectrometry. The analysis can be performed sequentially by feeding the liquid eluting from the liquid chromatography column directly to an ESI-MS setup.

The detection of ions is performed by a quadrupole mass spectrometer, sometimes referred to by the Agilent trade name "Mass Selective Detector" (MSD)

2.3 Crystallography

2.3.1 Theory

X-ray crystallography is the most powerful method to gain atomic resolution structures of biological molecules. The process involves purifying, crystallising and then collecting diffraction data from the crystal of the biological macromolecule in question. Electrons scatter X-rays to produce a continuous diffraction pattern of X-ray amplitude and phase. In a crystal, the diffracting electron density is convoluted with the lattice function, and consequently the pattern of diffracted X-rays is the product of the lattice transform, which itself is a lattice, the 'reciprocal lattice', and the transform of the density, which is consequently only sampled at discrete lattice points, defined by the integers h, k and l. The Fourier transform, equation 2.5, relates the electron density of a crystal structure and its diffraction pattern in reciprocal space as follows:

Equation 2.5

$$\rho(x,y,z) = \sum_{h}\sum_{k}\sum_{l} \underline{\mathbf{F}}(hkl) \exp\{-2\pi i (hx + ky + lz)\}$$

Where ρ(x,y,z) is the electron density and **F**(hkl) are the Fourier coefficients of the structure. **F** (hkl) is a complex quantity comprising a measurable amplitude, F (hkl), and non-directly observable phase, α (hkl).

The phases of the reflections are obtained, either experimentally or using molecular replacement methods to estimate the phase angles. Once information on the phase of the diffracted X-rays is available, the Fourier transform (Equation 2.1 above) can be used to construct the electron density in reciprocal space. A model of the protein molecule can then be built into the electron density using various geometric and other restraints from small molecule crystallography.

When X-rays are diffracted by a crystal they obey Bragg's law which is explained thus: Two beams with identical wavelength and phase approach a crystalline solid and are scattered off two different atoms within it. The lower beam (Figure 2.3.1) traverses an extra length of 2dsinθ. Constructive interference occurs when this length is equal to an integer multiple of the wavelength of the radiation giving equation 2.6.

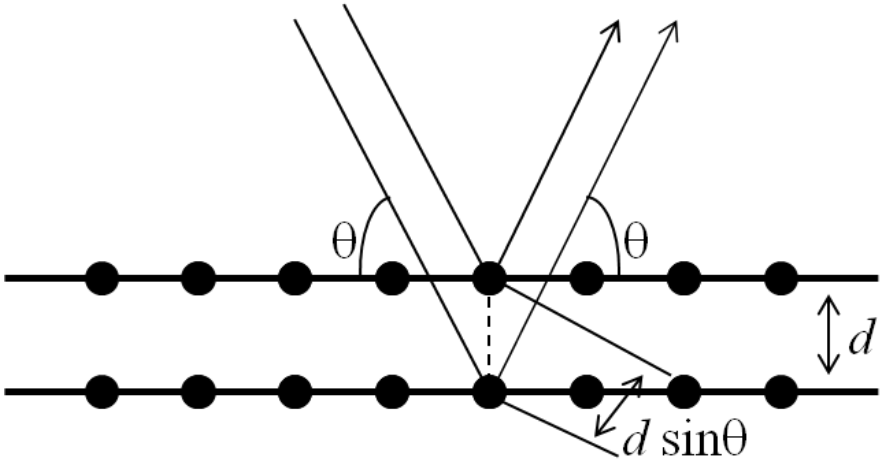


Figure 2.3.1 Explanation of Bragg Diffraction Crystalline particles within a lattice are shown as dots, equivalent planes within the crystal as horizontal thicker lines and incoming radiation as thin lines. The spacing, *d* is equal to the wavelength of the incoming radiation.

Equation 2.6 $n\lambda = 2d\sin\theta$

Where n is an integer, λ is the wavelength of incident X-ray radiation, d the spacing of interatomic planes and θ the angle of reflection.

2.3.2 Molecular Replacement and Structure Solution

The Molecular Replacement method is used to derive phase information from a partially-known existing model of the structure. Molecular Replacement (MR) also provides an initial starting model for refinement. Since there are an ever-increasing number of structures available in the PDB database, the chances are getting more and more reasonable that there will be at least a partial model available for your structure.

The first step in solving the phase problem by molecular replacement is to identify a suitable structure to use as a search model, this is usually via a blast search of the protein sequence against those of known protein structures. The method then involves finding the rotation and subsequently the translation that relates the search model to the unknown structure. Patterson superimposition methods are typically used in these calculations. The Patterson function, P(uvw), is calculated in the same way as electron density (equation 2.1) , but the structure factor amplitude is squared, $F(hkl)^2$, and the phase, α (hkl), is set to zero. The self-vectors needed for establishing the rotational relationship all lie in a volume around the origin of the Patterson function within a radius equal to the dimension of the molecule. For the rotation function the Patterson function calculated from the known structure can be rotated on to the Patterson function calculated from the experimental data with the maximum overlap giving the rotation matrix relating the known to the unknown structure. The translation function is then used to look for a correlation between cross-vectors of the Patterson function and therefore yield the position of the rotated search model in the unknown cell.

Phaser (McCoy, 2007) is an automated molecular replacement software package that combines anisotropy correction, likelihood enhanced fast rotation function, likelihood enhanced fast translation function, packing and refinement modes for multiple search models and allows a set of possible spacegroups to be searched in order to automatically solve a structure by molecular replacement.

However, if the correlation between search model and the actual protein structure in the crystal is low, the log likelihood gain (LLG) signal-to-noise of the search will below, there will be noise peaks and multiple ambiguous solutions. Signal-to-noise is judged using the Z-score, which is computed by comparing the LLG values from the rotation or translation search with

LLG values for a set of random rotations or translations. The mean and the RMS deviation from the mean are computed from the random set, then the Z-score for a search peak is defined as its LLG minus the mean, all divided by the RMS deviation, i.e. the number of standard deviations above (or below) the mean.

For a translation function the correct solution will generally have a Z-score (number of standard deviations above the mean value) over 5 and be well separated from the rest of the solutions. Table 2.1 gives a very rough guide to interpreting TF Z-scores.

TF Z-score	Have I solved it?
less than 5	no
5 - 6	unlikely
6 - 7	possibly
7 - 8	probably
more than 8*	definitely

Table 2.1 Explanation of TF Z-scores given by Phaser MR program

2.3.3 Model Building

Model building can be automatic using PHENIX (Adams *et al.* 2010) or ARPWARP (Langer *et al.* 2008) but sometimes manual intervention is needed to build poorer regions of the electron density map using graphics programs such as COOT (Emsle. *et al.,* 2010).

2.3.4 Structure refinement

Refinement is the optimization of the fit of the molecular model to the diffraction data. Maximum likelihood methods are employed in programs such as REFMAC (Winn *et al*., 2003) which adjusts the atomic positions and temperature factors to obtain a model that best accounts for the experimental data (maximizes the likelihood). Progress is measured by the fall in R-

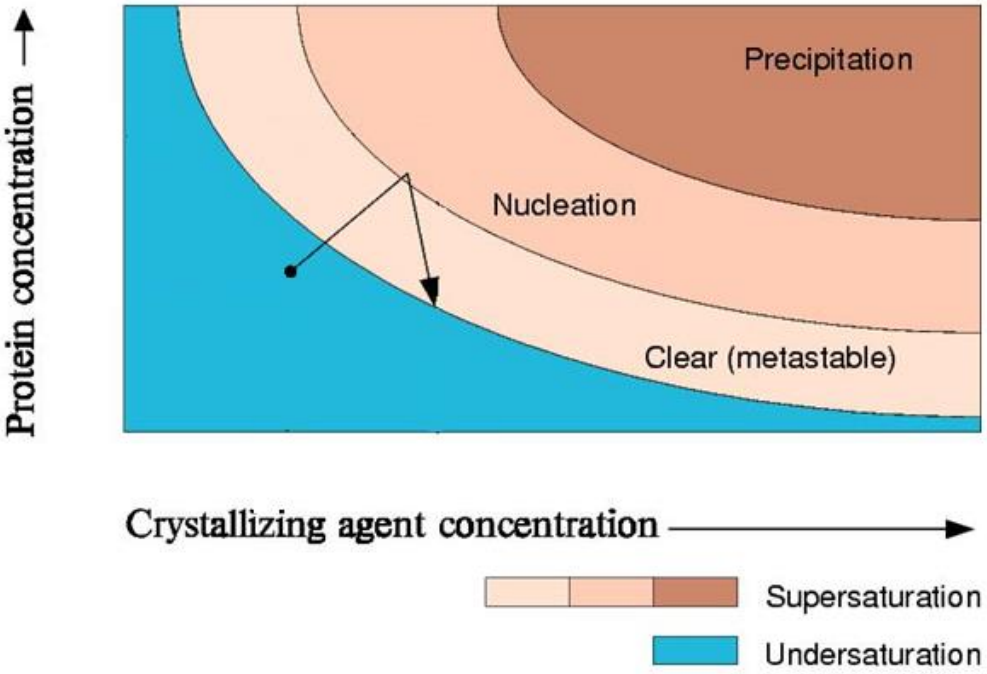
factor and R-free, as well as by the likelihood scores. In reality, when optimizing a model at medium resolution such as 2.0–3.0 Å, cycles of refinement followed by manual rebuilding are used until a good agreement is seen between electron density map and model and the refined residuals and stereochemical quality indicators, such as deviation in bond lengths, are adequate.

2.3.5 Validation and deposition

Structures are validated against observations not included in the refinement process, such as agreement of main-chain torsion angles with allowed regions of the Ramachandran plot or of side-chains rotomers with accepted rotomer distributions. Programs such as PROCHECK (Laskowski *et al.*, 1993) or MolProbity are used to validate structures and protein databank deposition imposes a reasonable stringency in checking the deposited structure.

2.3.6 Crystalisation

Crystallization conditions were screened using hanging drop vapour diffusion. Crystalisation as a process first involves the nucleation of molecules in a saturated solution (Figure 2.3). Initially this requires the protein to be undersaturated and soluble. Then, as the precipitant reservoir draws water from the protein drop, the concentration of protein increases allowing nucleation and the growth of a crystal. However, if crystals begin to grow, the concentration of protein in solution will decrease.



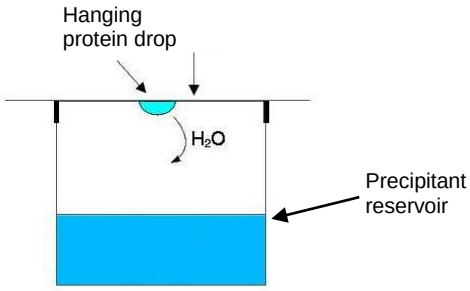


Figure 2.3.2 Top: Depiction of concentration changes necessary for crystal growth As solvent in the hanging drop evapourates, the concentration of protein in the drop increases, causing the solution to go from undersaturated to a concentration where nucleation can occur. As protein forms nuclei, the concentration in solution decreases allowing metastable crystal growth.

Bottom: Cartoon of Hanging Drop Set up A cartoon showing a typical set-up for hanging drop. Curly arrow shows net movement of water and dehydration of protein drop.

Crystallisation conditions were screened using hanging drop vapour diffusion (Figure 2.3.2) in 96 well plates wherein protein and precipitant drops were in the range of 200-400nL. These were set up set up using a nanodrop Mosquito liquid handling robot (TTP Labtech).

Sparse matrix screens such as JCSG+ / PACT Premier (Molecular Dimensions) were used to trial as many different conditions of pH, ionic strength and precipitating agent as possible.

2.3.7 Crystalisation optimisation

However, scaling up, to 1-2µL drops in 24 well plates to make crystal handling easier proved unsuccessful by vapour diffusion alone. Further crystals were however successfully obtained by whisker streak seeding, using the initial hit as a source of seeds. Selected crystals were removed from the mother liquor and transferred to a solution containing mother liquor plus 5% propane-1,2-diol as a cryoprotectant and flash cooled in liquid nitrogen.

2.4 NMR spectroscopy

2.4.1 Theory

Both X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy have their limitations and advantages. Protein NMR is limited by the size of the protein as data is collected about individual atoms with very similar, but subtly different properties, the major difficulty become assigning which chemicals shifts correspond to which atoms in the protein, and as the size of the protein increase so does the number of atoms and the difficulty is confounded. NMR is very sensitive to the chemical environment of individual atoms and as such is a powerful method to explore protein/ligand interactions and protein dynamics.

The phenomenon of NMR arises because subatomic particles (electrons, protons and neutrons) can be imagined as spinning on their axes. In many atoms (such as ^{12}C) these spins are paired against each other, such that the nucleus of the atom has no overall spin. However, in atoms with an odd number of neutrons and/or protons the nucleus does possess an overall spin. For example, in atoms such as ^{1}H and ^{13}C which both have I= ½ spin states

A nucleus with spin 1/2 will have 2 possible orientations. In the absence of an external magnetic field, these orientations are of equal energy. If a magnetic field is applied, then the energy levels split. Each level is given a *magnetic quantum number, m*.

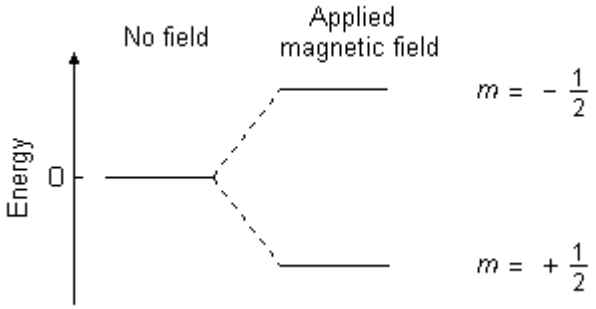


Figure 2.4 Energy levels for a nucleus with spin quantum number I=½ A diagram depicting the 2 possible energy states of an I=1/2 nucleus under an applied magnetic field. Lower energy state when aligned with and higher when aligned against an applied magnetic field (B_0)

The most common natural occurring isotopes of the atoms most relevant to biomolecules (12C, 14N) have I=0 and as a result are 'NMR silent'. However, it is possible to produce proteins containing isotopes which do provide NMR signals (13C, 15N) and this process requires labelling protein. IN essence this is a straight forward process and involves producing the protein recombinantly and growing the protein producing organism in a broth / environment supplemented with glucose and nitrogen containing this heavier isotopes.

15N singly or 13C- and15N-doubly isotopically labeled protein were used to record various spectra. Protein used to record 3D experiments was around 0.5mM. Sample concentrations ranging from 25μM to 0.5mM were used to record 2D spectra depending on the nature of the experiment and the time available.

2.4.2 Spectra

1H-15N HSQC

1H-15N Heteronuclear Single Quantum Coherence (HSQC) spectra and 1H-15N Sofast-HMQC were used for purposes such as sample evaluation, thermal melt analysis, titration experiments and for assignment.

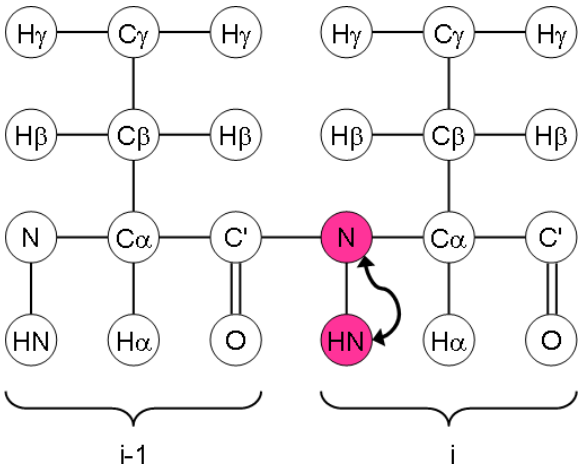


Figure 2.5 Diagram depicting the magnetisation transfer in a typical 1H-15N HSQC experiment.

In an HSQC experiment, magnetization is transferred from hydrogens to attached to ^{15}N nuclei via J-coupling (Figure 2.5). The chemical shift is evolved on the nitrogen and the magnetisation is then transferred back to the hydrogen for detection. This shows all H-N correlations.

These amides are predominantly the backbone amide groups, but Trp side-chain Nε-Hε groups and Asn/Gln side-chain Nδ-Hδ2/Nε-Hε2 groups are also visible. The Arg Nε-Hε peaks are in principle also visible, but because the Nε chemical shift is outside the region usually recorded, the peaks are folded/aliased (this essentially means that they appear as negative peaks and the Nε chemical shift has to be specially calculated). If working at low pH the Arg Nη-Hη and Lys Nζ-Hζ groups can also be visible, but are also folded/aliased.

The spectrum is rather often termed the "fingerprint" of a protein as it is highly individual to a particular protein. It was the first heteronuclear experiment performed and used to assess whether the protein was folded In solution and whether further NMR experiments were likely to succeed.

HNCACB

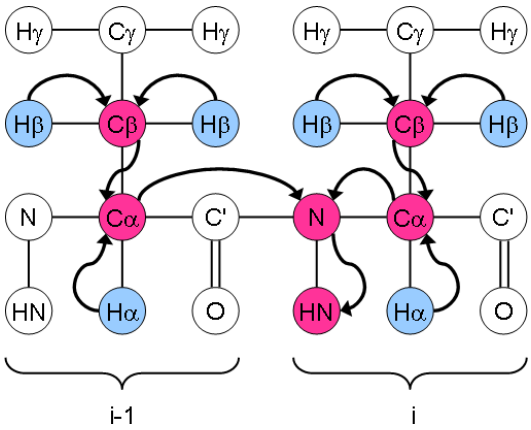


Figure 2.6 Diagram depicting the magnetisation transfer in a typical HNCACB experiment.

Magnetisation is transferred from Hα and Hβ to Cα and Cβ, respectively, and then from Cβ to Cα. From here it is transferred first to NH and then to HN for detection. Transfer form Cαi-1 can occur both to N_{i-1} and N_i, or viewed the other way, magnetisation is transferred to N_i from both $C\alpha_i$ and $C\alpha_{i-1}$. Thus for each NH group there are two Cα and Cβ peaks visible. The chemical shift is evolved simultaneously on Cα and Cβ, so these appear in one dimension. The chemical shifts evolved in the other two dimensions are 15NH and 1HN.

CBCA(co)NH

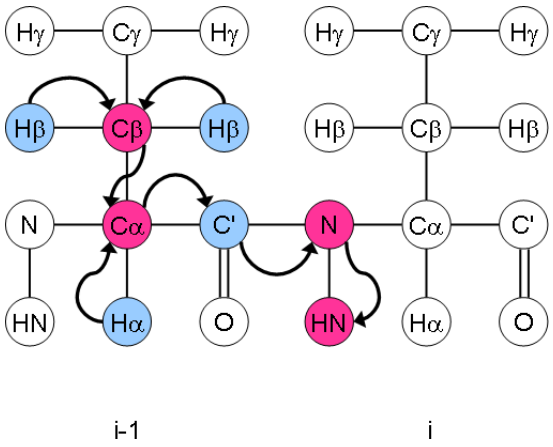


Figure 2.7 Diagram depicting the magnetisation transfer in a typical CBCA(co)NH experiment.

In recording a CBCA(co)NH spectrum, magnetisation is transferred from Hα and Hβ to Cα and Cβ, respectively, and then from Cβ to Cα. From here it is transferred first to CO, then to NH and then to HN for detection (Figure 2.7). The chemical shift is evolved simultaneously on Cα and Cβ, so these appear in one dimension. The chemical shifts evolved in the other two dimensions are NH and HN. The chemical shift is not evolved on CO.

Nuclear Overhauser Effect (NOE)

The Nuclear Overhauser Effect (NOE) is a phenomenon observed by cross relaxation between dipolar coupled spin systems (Figure 2.8). It is exceptionally important tool for structural studies.

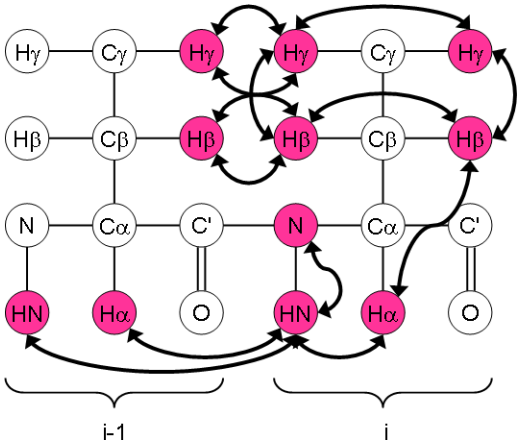


Figure 2.8 Diagram depicting the magnetisation transfer in a typical CBCA(co)NH experiment.

2.4.3 NMR titration

NMR titration is achieved by recording a 1H-15N HSQC/ 1H-15N Sofast-HMQC spectra of the labeled protein. Cross titrations were performed to eliminate the dilution of the labeled protein. In cross titration, two samples were prepared; one sample was labeled protein only and the second sample had the same concentration of labeled protein and a high equivalent (i.e. 10 equivalents) of unlabeled titration partner. During the titraion the concentration of labeled protein did not change, so it was not necessary to correct for dilution.

2.5 Protein Structural Simulation and Docking

The program HADDOCK is a protein-protein docking algorithm which uses biochemical and biophysical information to provide a model of the protein-protein complex (Dominguez, Boelens et al. 2003). It uses experimental data (i.e. chemical shift perturbation data, site specific mutagenesis data). The information on interacting residues is introduced as ambiguous interaction restrains (AIRs) to drive the docking.

In the case of NMR titration data, the active residues correspond to all residues showing a significant chemical shift perturbation upon complex formation as well as high solvent accessibility in the free form of protein. The passive residues correspond to the residues that show a less significant chemical shift perturbation and/or that are surface neighbours of the active residues with high solvent accessibility (>50%). The AIRs files are generated by entering active and passive residue on http://www.nmr.chem.uu.nl/haddock/. 1000 docking structures were generated for each calculation. The output structures are ranked according to their intermolecular energy.

Chapter 3: Methods & Protocols

Molecular & Cellular Biology

3.1 LB Media

LB Media used for growing *E. coli* was made up using standard protocol (below) and made up to 1L and sterilized by autoclaving at 121°C for 15 min.

Component	Concentration (g/L)
Bacto-tryptone (A pancreatic digest of casein)	10
Yeast extract	5
Sodium chloride	5

Table 3.1 Contents of LB media

3.2 Minimal (M9) Media - 10x Stock

Component	Concentration (g/L)
Na_2HPO_4	60
KH_2PO_4	30
Sodium chloride	5

Table 3.2 Contents of M9 media

3.3 BL21(DE3) Transformation

1. Remove cells from –80°C and let thaw on wet ice.

2. Gently mix cells by lightly flicking tube. Aliquot ~50-100µl of cells into chilled, 17 x 100mm polypropylene tube(s), e.g., Falcon 2059. Unused cells may be refrozen, but a small drop in efficiency may result. For optimal recovery, refreeze cells in a dry ice/ ethanol bath prior to -80°C storage.

3. Add DNA solution (≤5µl per 50µl cells) to cell suspension and gently swirl tube(s) for a few seconds to mix. If a control is desired, repeat this step with 2µl of the provided Control Vector (pUC19) in a separate tube.

4. Incubate on ice for 30 minutes.

5. Place tube(s) in 42°C water bath for ~30 to 45 seconds without shaking.

6. Replace tube(s) on ice for ~2 minutes.

7. Dilute transformation reaction(s) to 1ml by addition of 900-950µl LB media.

8. Shake tube(s) ~200 rpm for 60 minutes at 37°C.

9. Plate by spreading 5-200µl of cell transformation mixture on LB agar plates containing appropriate antibiotic and incubate overnight at 37°C.

3.4 Mutagenesis – Stratagene Quickchange Protocol

a) Mutant Strand Synthesis Reaction (Thermal Cycling)

1. Prepare the ds-DNA template either by standard miniprep protocols (e.g. StrataPrep Plasmid Miniprep Kit, catalog #400761) or by cesium chloride gradient purification.

2. Prepare the mutant strand synthesis reactions for thermal cycling as indicated below. Add the components in the order listed then mix gently by pipetting or tapping the reaction tube.

Reaction Component	**Templates <5kb**	**Templates >5kb**
10x QuickChange Lightning Multi reaction buffer	2.5 µL	2.5 µL
Double distilled H_2O	To final volume of 25 µL	To final volume of 25 µL
ds-DNA template	50 ng	100 ng
Mutageneic primers	100 ng each primer for 1-3 primers; 50 ng each primer for 4-5 primers	100 ng each primer for 1-3 primers; 50 ng each primer for 4-5 primers
dNTP Mix	1 µL	1 µL
QuickChange Lightning Multi enzyme blend	1 µL	1 µL

Table 3.3 Reaction components for Quick Change Lightening II mutagenesis reactions

3. Cycle the reactions using the cycling parameters outlined in the table below. (For the control reaction, use a 2.5-minute extension time.)

4. Following temperature cycling, place the reaction on ice for 2 minutes to cool the reaction to ≤37°C.

b) Dpn I Digestion of the Amplification Products

1. Add 1 µl of Dpn I restriction enzyme directly to each amplification reaction.

2. Gently and thoroughly mix each reaction mixture by pipetting the solution up and down several times. Spin down the reaction mixtures in a microcentrifuge for 1 minute, then immediately incubate each reaction at 37°C for 5 minutes to digest the parental (nonmutated) ds-DNA.

c) Transformation of XL10-Gold Ultracompetent cells

1. Gently thaw the XL10-Gold ultracompetent cells on ice. For each mutagenesis reaction, aliquot 45 µl of the ultracompetent cells to a prechilled 14-ml BD Falcon polypropylene round-bottom tube.

2. Add 2 µl of the β-ME mix provided with the kit to the 45 µl of cells. Using an alternative source of β-ME may reduce transformation efficiency.

3. Swirl the contents of the tube gently. Incubate the cells on ice for 10 minutes, swirling gently every 2 minutes.

4. Transfer 1.5 µl of the Dpn I-treated DNA from each mutagenesis reaction to a separate aliquot of the ultracompetent cells. Optional Verify the transformation efficiency of the XL10-Gold ultracompetent cells by adding 1 µl of 0.01 ng/µl pUC18 control plasmid (dilute the provided pUC18 DNA 1:10) to another 45-µl aliquot of cells.

5. Swirl the transformation reactions gently to mix, then incubate the reactions on ice for 30 minutes.

6. Preheat NZY+ broth (see Preparation of Media and Reagents) in a 42°C water bath for use in step 9. Note Transformation of XL10-Gold ultracompetent cells has been optimized using NZY+ broth.

7. Heat-pulse the tubes in a 42°C water bath for 30 seconds. The duration of the heat pulse is critical for obtaining the highest efficiencies. Do not exceed 42°C. Note This heat pulse has been optimized for transformation in 14-ml BD Falcon polypropylene round-bottom tubes.

8. Incubate the tubes on ice for 2 minutes.

9. Add 0.5 ml of preheated (42°C) NZY+ broth to each tube and incubate the tubes at 37°C for 1 hour with shaking at 225–250 rpm.

10. Plate 100 µL of each transformation reaction, on agar plates containing the appropriate antibiotic for the plasmid vector. For the mutagenesis and transformation controls, spread cells on LB–ampicillin agar plates (see Preparation of Media and

Reagents) that have been prepared with 80 µg/ml X-gal and 20 mM IPTG (see Preparing the Agar Plates for Color Screening).

3.5 Expression of OutB

After successful transformation, individual colonies were used to inoculate 10ml "starter cultures" containing 100µgml^{-1} ampicillin and were grown at 37°C for 16h, typically overnight. These were used to inoculate 1L cultures of LB media. Cells were grown to an OD_{600} of 0.6–0.8 and expression induced by addition of 0.2 mM IPTG. Induced protein expression was allowed to proceed for 16 h at 20°C typically overnight.

3.6 Expression of GspD^{OutD} N0

After successful transformation, individual colonies were used to inoculate 10ml "starter cultures" containing 100µgml^{-1} ampicillin. The culture was grown at 37°C and induced with 1mM IPTG when optical density (OD600) reached 0.6-0.8. Cells culture continued to be incubated at 37°C for 4 h post induction.

3.7 GST-Tag Purification of OutB

Cells from a 1L culture were harvested by centrifugation at 10,000g for 20min then resuspended in 20ml of cell culture of phosphate buffered saline (PBS) containing 0.01M phosphate buffer, 0.0027M KCl and 0.137M NaCl, pH 7.4. It is possible to store resuspended cell pellets at this stage by freezing at -80°C. Subsequent steps can be completed after thawing the cell pellet.

Cells were lysed on ice by sonication at 20W with 20 replicates of 10s sonication followed by 20s of cooling.. Cell debris was removed by centrifugation at 35,000g for 20min.Afterwards, the supernatant was decanted and incubated with 2.5ml of Glutathione sepharose beads (G.E. Healthcare) for 6h. The glutathione sepharose was then collected in a column and washed with 20 column volumes (50 ml) of PBS.

On column cleavage of the GST tag required washing Glutathione sepharose with bound GST-OutB in PreScission protease (G.E. Healthcare) buffer containing (50mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.0) and addition of 20µL of PreScission protease to 5ml bead volume of glutathione sepharose. Cleavage was at 4°C for 6hr. Elutate containing the tag-less $GspB^{112-220}$ was then run on a Superdex S200 gel filtration column and SDS-PAGE used to assess and confirm purity prior to crystallisation

3.8 His-tag purification of $GspD^{OutD}$ N0

GspD was expressed with a C-terminal hexahistidine tagged and purified as follows. Cells were collected and the pellet was lysed by sonication as detailed above in a binding buffer (20mMTris, 400mM NaCl, 20mM imidazole pH8.0). The soluble protein fraction was then separated by centrifugation at 15,000g for 15 mins and applied to the pre-packed His-Trap column (GE Healthcare). The column was pre-equilibrated in binding buffer, using a flow rate of 1ml/min. The non-specifically binding proteins were eliminated by washing with 20 column volumes (CV) of Ni column washing buffer (20mM Tris, 400mM NaCl, 20mM imidazole pH 6.5) until no more protein was detected in the flow through as determined using Bradford solution.

The bound, target protein was eluted using elution buffer (20mM Tris, 400mM NaCl, 500mM imidazole, pH6.5) again until no further protein was detected in flow through, tested using Bradford solution. The eluted fractions were checked using SDS-PAGE.

3.9 Isotope Labelling for NMR

Recombinant uniformly 15N- 13C- labeled full length OutB (or OutD-N0) was

expressed using the relevant vector in BL21(DE3) (Novagen),. Starter cultures were grown in LB media, and used to innoculate growth in M9 minimal media.

M9 minimal media supplemented with trace elemnts and isotope sources as listed below.

All minimal media was 0.2μm filter sterilized.

Compound	Concentration (g/L)
Na_2HPO4	6
KH2PO4	2
NaCl	0.5
MgSO4	0.25
CaCl2	0.015
FeSO4	0.015
Thiamine	0.001
Biotin	0.001
$^{15}NH_4Cl$	1
^{13}C D-glucose	2
Antibiotic – Ampicillin	0.1

Table 3.4 Contents of isotopically enriched M9 media with supplementary trace elements

3.10 SDS PAGE

SDS-PAGE gels were cast and run using the Bio-Rad mini tank system.

Thickness (mm)	Volume of Stacking Gel (ml)	Volume of Resolving Gel (ml)
0.75	2	4
1.0	3	6
1.5	4	8

Table 3.5 Volumes required for casting gels in Bio-Rad system

Stacking gels are always of a consistent percentage acrylamide content in order to ensure movement of the protein mixture into the resolving gel.

Component	Volume (ml)
H_2O	2.975
0.5 M Tris-HCl, pH 6.8	1.25
10% (w/v) SDS	0.05
Acrylamide/Bis-acrylamide (30% / 0.8% w/v)	0.67
10% (w/v) ammonium persulfate (AP)	0.05
TEMED	0.005

Table 3.6 Component volumes for 4ml stacking gel

Resolving gels consisted of crosslinked polyacrylamide made up at varying concentrations in order to allow resolution of various molecular weight proteins.

Acylamide percentage	**6%**	**8%**	**10%**	**12%**	**15%**
H_2O	5.2ml	4.6ml	3.8ml	3.2ml	2.2ml
Acrylamide/Bis-acrylamide	2ml	2.6ml	3.4ml	4ml	5ml
1.5M Tris(pH=8.8)	2.6ml	2.6ml	2.6ml	2.6ml	2.6ml
10% (w/v)SDS	0.1ml	0.1ml	0.1ml	0.1ml	0.1ml
10% (w/v) ammonium persulfate (AP)	100µl	100µl	100µl	100µl	100µl
TEMED	10µl	10µl	10µl	10µl	10µl

Table 3.7 Component volumes for 5ml stacking gel at various acrylamide concentrations

Samples were boiled in a buffer to ensure denaturation of the protein and addition of SDS to ensure charge of the protein. DTT or β-ME can be excluded if you wish to run proteins under non-reducing conditions (i.e. if disulphide bonds may be present).

10% w/v	SDS
10 mM	Dithiothreitol, or beta-mercapto-ethanol
20 % v/v	Glycerol
0.2 M	Tris-HCl, pH 6.8
0.05% w/v	Bromophenolblue

Table 3.8 5x Sample Buffer composition

1x Running Buffer

25 mM	Tris-HCl
200 mM	Glycine
0.1% (w/v)	SDS

Table 3.9 Running buffer composition

Protocol:

1. Make the separating gel:

    Set the casting frames (clamp two glass plates in the casting frames) on the casting stands.

    Prepare the gel solution (as described above) in a separate small beaker.

    Swirl the solution gently but thoroughly.

    Pipet appropriate amount of separating gel solution (listed above) into the gap between the glass plates.

    To make the top of the separating gel be horizontal, fill in water (either isopropanol) into the gap until a overflow.

    Wait for 20-30min to let it gelate.

    Make the stacking gel:

    Discard the water and you can see separating gel left.

    Pipet in stacking gel untill a overflow.

    Insert the well-forming comb without trapping air under the teeth. Wait for 20-30min to let it gelate.

2. Make sure the stacking gel is completely set and remove the well comb. Take the glass plates out of the casting frame and set them in the running buffer dam. Pour the running buffer into the inner chamber and keep pouring after overflow until the buffer surface reaches the required level in the outer chamber.

3. Prepare the samples:

    Mix samples with sample buffer

    Heat (90°C) for 5-10 min.

4. Load prepared samples into wells and make sure not to overflow. Load protein marker into the first lane.

5. Attach top and connect the anodes.

5. Set an appropriate voltage and run the electrophoresis.

Total running time: stop SDS-PAGE running when the bromophenol blue dye front almost reaches the foot line of the glass plate. Generally, about 1 hour for a 120V voltage and a 12%

separating gel. For a separating gel possessing higher percentage of acrylamide, the time will be longer.

3.11 Circular Dichroism

Far-UV CD measurements were made using a Chiroscan spectropolarimeter equipped with temperature controller. Spectra were recorded in 20mM Tris at pH7.0 using 1 mm path length fused silica cuvettes. Protein samples used for CD were at a concentration of 0.05mg/ml. The spectra are presented as differential absorbance after baseline subtraction.

3.12 Dynamic Light Scattering

Dynamic light scattering was measured using a DynaPro Molecular sizing instrument running DYNAMICS V6 software. The quartz cuvettes (45 µl) were washed 5 times each with 1% Triton/H_2O then with dd H_2O and dried with compressed nitrogen gas. The exterior surface of cuvette was cleaned by wiping with ethanol and lens tissue. The protein concentration used was 4mg/ml (or as stated). Each sample was passed through a 0.2µm filter and a minimum of 20 measurements were recorded at 25°C.

3.13 Mass Spectrometry

To prepare samples, formic acid was mixed with the protein sample to a final concentration of 1% (v/v). The sample was first bound to a C5 reverse-phase HPLC column and then eluted with an acetonitrile gradient. The eluted fractions from the HPLC column were assessed by electrospray ionisation mass spectroscopy.

3.14 NMR Methods

3.14.1 Sample preparation

Main-chain amide resonance assignments required the spectra detailed in Table 3.10 to be collected

Experiment	Purpose of experiment
^{1}H NMR	Sample evaluation
^{1}H-^{15}N HSQC	Reference spectrum
^{1}H-^{15}N HMQC	Titration
HNCACB CBCA(CO)NH HNCO HNCA	Backbone assignment
CBCAHA HACA(CO)N HCCH-TOCSY	Side-chain assignment

Table 3.10 NMR Spectra required for backbone assignment

All NMR spectra were acquired at 37 ºC on a Bruker Avance spectrometer operating at 600 MHz for 1H nuclei using a 5 mm inverse detection triple-resonance z-gradient probe.

Typically, 1D 1H NMR spectra were obtained with 128 transients, consisting of 64 K complex points over a sweep width of 8,389 Hz. Data was processed using an exponential Fourier transform window, with a 2.0 Hz line broadening. Residual water was suppressed using either pre-saturation, excitation sculpting with gradients solvent suppression or the Watergate method (Liu, M., et al. 1998).

Phase sensitive 2D 1H-15N HSQC NMR spectra were acquired using echo-antiecho gradient selection. The 1H acquisition parameters were 0.122 seconds acquisition time, 1 second fixed

delay and 2048 complex (t2) points while 256 complex points were collected for the 15N dimension. Thirty two transients were recorded for each spin lattice relaxation time (t1) interval. The 1H and 15N dimensions possessed spectral widths of 14 ppm and 25 ppm respectively. The 15N dimension was zero filled to 256 data points before squared cosine apodisation and Fourier transformation.

3.14.2 Data Processing, spectral analysis and resonance assignment

Standard triple resonance backbone assignment of protein using HNCACB and CBCACONH spectra was used (Figure 3.1). The idea is that the HNCACB spectra correlates each NH group with the chemical shift of the current residue's CA and CB (strongly) and that of neighbouring residue's CA and CBs (weakly). The CBCA(CO)NH spectra only correlates the NH group of the current residue to the preceding CA and CB chemical shifts. By comparing both these spectra, the current residue and its preceding CA CB resonances can be identified. The succeeding amide proton chemical shift was identified using the chemical shifts of the preceding CA CB to search for the amide giving same chemical shift within the current residue's CA, CB on HNCACB spectra.

3.14.3 NMR Titration

NMR titration is achieved by recording a 1H-15N HSQC/ 1H-15N Sofast-HMQC spectra of the labeled protein. In serial titrations, cross titrations were performed to eliminate the dilution of the labeled protein. In cross titration, two samples were prepared; one sample was labeled protein only and the second sample had the same concentration of labeled protein and a high equivalent (i.e. 10 equivalents) of unlabeled titration partner. During the titraion the concentration of labeled protein did not change, so it was not necessary to correct for dilution.

Chapter 4: Crystal Structure of OutB

Aim:
OutB is known from *in vivo* studies to be highly important for the assembly of the type 2 secretion system in several Gram negative bacteria. The aim of this Chapter is to determine the structure of the GspB homologue from *Dickeya dadantii*, OutB, so that its role in assembly of the secretion system can be probed at the molecular level.

4.1 Introduction

GspB homologues have been described in at least three genera of bacteria, Klebsiella, Aeromonas, and the Erwinia complex (including *Dikeya dadantii*). However, their requirement in the secretion process seems to be different. In Aeromonas and Vibrio species, GspB is associated with another auxiliary component GspA. GspA is known to interact with peptidoglycan (Sandkvist *et al.* 2011). As a result of this association of a peptidoglycan binding protein and GspB, it has been hypothesized that GspB is involved in assembly of the secretin possibly by reorganization of the peptidoglycan network or in combination with the inner membrane platform.

Dickeya species lack any detectable GspA homologue, so in these bacteria GspB is believed to act alone. Previous work published by Shevchik and Condemine 2001, indicated that the GspB homologue OutB, from *Dickeya dadantti* (formerly *Erwinia chrysanthemi*), is an inner membrane protein with a short cytoplasmic tail, transmembrane helix (figure 4.1) and large periplasmic domain.

However, fractionation experiments indicate that it could be associated with the outer membrane through its C-terminal part. The secretion deficiency of an *E. chrysanthemi*

outB mutant can be reversed by the addition of an inducer of the *kdgR* regulon which results from the increased expression of the secretin OutD. Also, secretion can be restored in an *outB* negative mutant by introducing the *outD* gene on a plasmid.

Several experiments suggest an interaction between OutB and OutD. In *D. dadantii*, the presence of OutD stabilizes OutB. OutD expressed in *Escherichia coli* can be protected from proteolytic degradation by the co-expression of OutB. This effect does not require the N-terminal, transmembrane segment of *outB*. OutB can be cross-linked with OutD by formaldehyde. These results indicate that OutB could act with OutD in the functioning of the Out secretion machinery.

4.2 Bioinformatic Analyses

The major functional domain of OutB is presumed to lie at the C-terminus. This putative domain is approximate 170 residues long and resides in the periplasm. Alignment of full length GspB protein sequences (figure 4.2) from various species, shows that the transmembrane helix (Figure 4.1) and C-terminal domains are the most highly conserved regions across all species. Dickeya species appear to have a C-terminal extension of approximately 15 residues which is itself highly conserved across members of this genera, but not across GspB homologues.

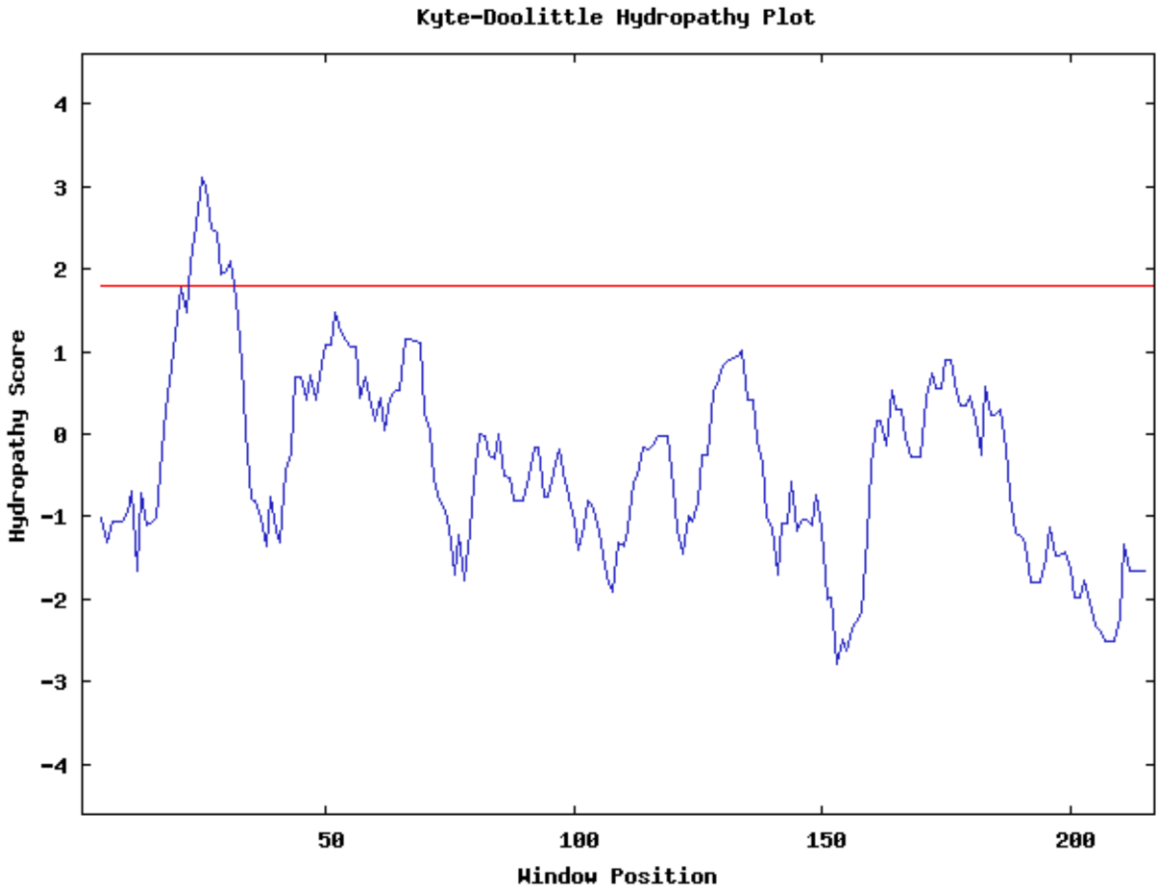


Figure 4.1 Hydropathy plot of the OutB sequence of *Dickeya dadantii* residue number is indicated on x-axis and hydropathy score on the y-axis. Note the peak between residues 20-40 indicating the transmembrane helix.

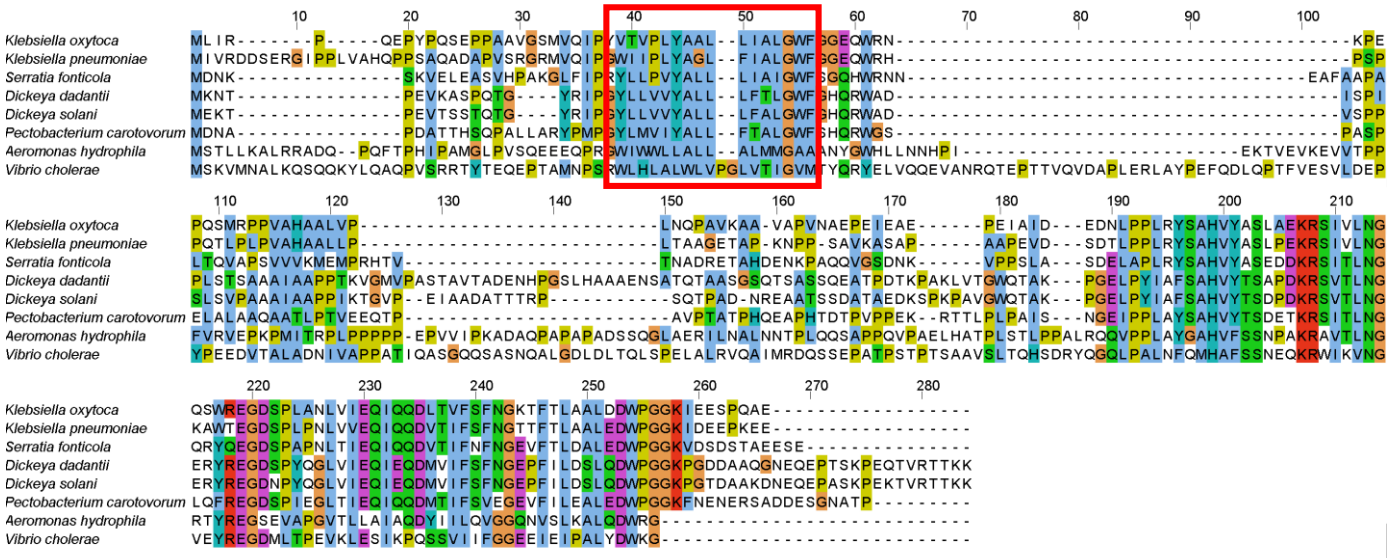


Figure 4.2 Sequence alignment of full length GspB proteins from 8 species with the N-terminal transmembrane helix (TMH) highlighted in red. The numbering above is for reference only, and does not refer to one particular amino acid sequence.

Given that OutB possess only a short cytoplasmic tail and single transmembrane helix, it is safe to assume the major functional domain is the highly conserved C-terminal domain. Therefore efforts to solve the structure dealt only with the soluble C-terminal domain rather than face the difficulties of dealing with a full length membrane protein, the structure of which may reveal few if any insights into the functioning of the protein.

4.3 Expression and Purification of OutB

DNA encoding the highly conserved region of the periplasmic domain of *D. dadantii* GspB$^{112-220}$, residues 112 to 220, was subcloned into the pGEM-T Easy (Promega) vector, its sequence confirmed and then the insert cloned into the pGEX-6P-3 vector encoding a cleavable N-terminal Glutathione S-transferase (GST) affinity tag. This construct was transformed in to *E. coli* strain BL21(DE3) for protein expression. 10ml inoculation cultures containing 100µgml^{-1} ampicillin were grown at 37°C for 16h typically overnight. These were used to inoculate 1L cultures of Lysogeny broth (LB) media. Cells were grown to an OD_{600} of 0.6 – 0.8 and expression induced by addition of 0.2mM isopropyl β-D-1-thiogalactopryanoside (IPTG). Induced protein expression was allowed to proceed for 16h at 20°C typically overnight.

Cells from a 1L culture were harvested by centrifugation at 10,000g for 20min then resuspended in 20ml of cell culture of phosphate buffered saline (PBS) containing 0.01M phosphate buffer, 0.0027M KCl and 0.137M NaCl, pH 7.4 and lysed on ice by sonication at 20W with 20 replicates of 10s sonication followed by 20s of cooling. Cell debris was removed by centrifugation at 35,000g for 20min and the supernatant was incubated with 2.5ml of Glutathione sepharose beads (G.E. Healthcare) for 6h. The glutathione sepharose was then collected in a column and washed with 50ml of PBS.

On column cleavage of the GST tag required washing Glutathione sepharose with bound GST-OutB in PreScission protease (G.E. Healthcare) buffer containing (50mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 7.0) and addition of 20µL of PreScission protease to 5ml bead volume of glutathione sepharose. Cleavage was at 4°C for 6hr Figure 4.3). Elutate containing the tag-less GspB$^{112-220}$ was then run on a Superdex S200 gel filtration column and SDS-PAGE used to assess and confirm purity prior to crystallisation trials.

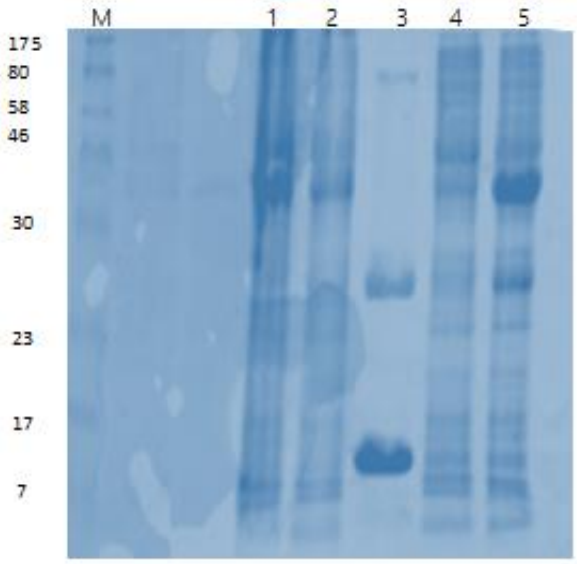


Figure 4.3 SDS-PAGE gel showing expression of OutB112-220. Lanes from left to right are:

M Molecular marker, masses in kDa

1. Lysed cells
2. Supernatant
3. Post digest of GST-Tagged outB with PreScission protease
4. GST column flow through
5. Post induction with 0.5mM IPTG

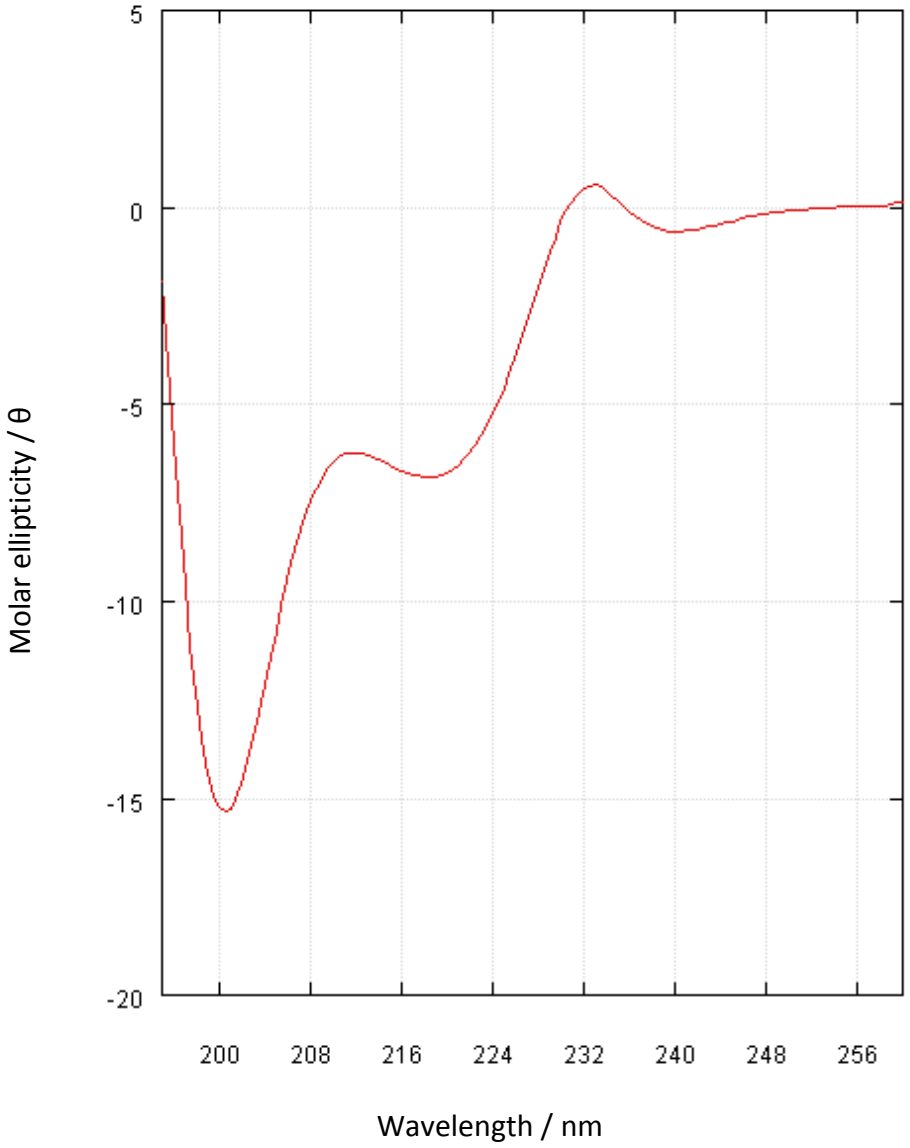


Figure 4.4 CD Spectra of OutB 112-220
Circular dichroism spectra of purified OutB (residues 112-220) recorded at 10°C in 20mM tris pH7.0. The spectrum contains 2 distinct negative peaks at 200nm and 220nm and a smaller positive peak at 235nm.

4.4 Circular dichroism (CD) charecterisation of the periplasmic domain of OutB

Having confirmed the plasmid construct successfully produced protein and the purification protocol allowed for the isolation of uncontaminated protein, it was necessary to characterise residues 112-220 of OutB to confirm a folded, stable domain was formed in solution.

The CD spectra of OutB112-220 (Figure 4.4) shows distinct peaks, different to that which would be expected from an unstructured polypeptide forming a random coil in solution. This indicates that the domain formed by residues 112-220 forms a folded, 3D, tertiary structure in solution. The peak present at 232nm may indicate the possibility of extended poly-proline II like extended helix. Further analysis was conducted in order to assess the stability of the OutB constructed which had been successfully expressed and purified.

Thermal melt CD analysis was conducted, whereby CD spectra are recorded at ever increasing temperatures further supports this analysis. The peaks present in the CD spectra at 10°C indicated the existence of secondary structure elements; that were lost at temperatures greater than 30°C (Figure 4.5). This is a typical stable temperature range for *in vitro* purified protein. Above 30°C, the CD trace was highly characteristic of random coil, unfolded protein.

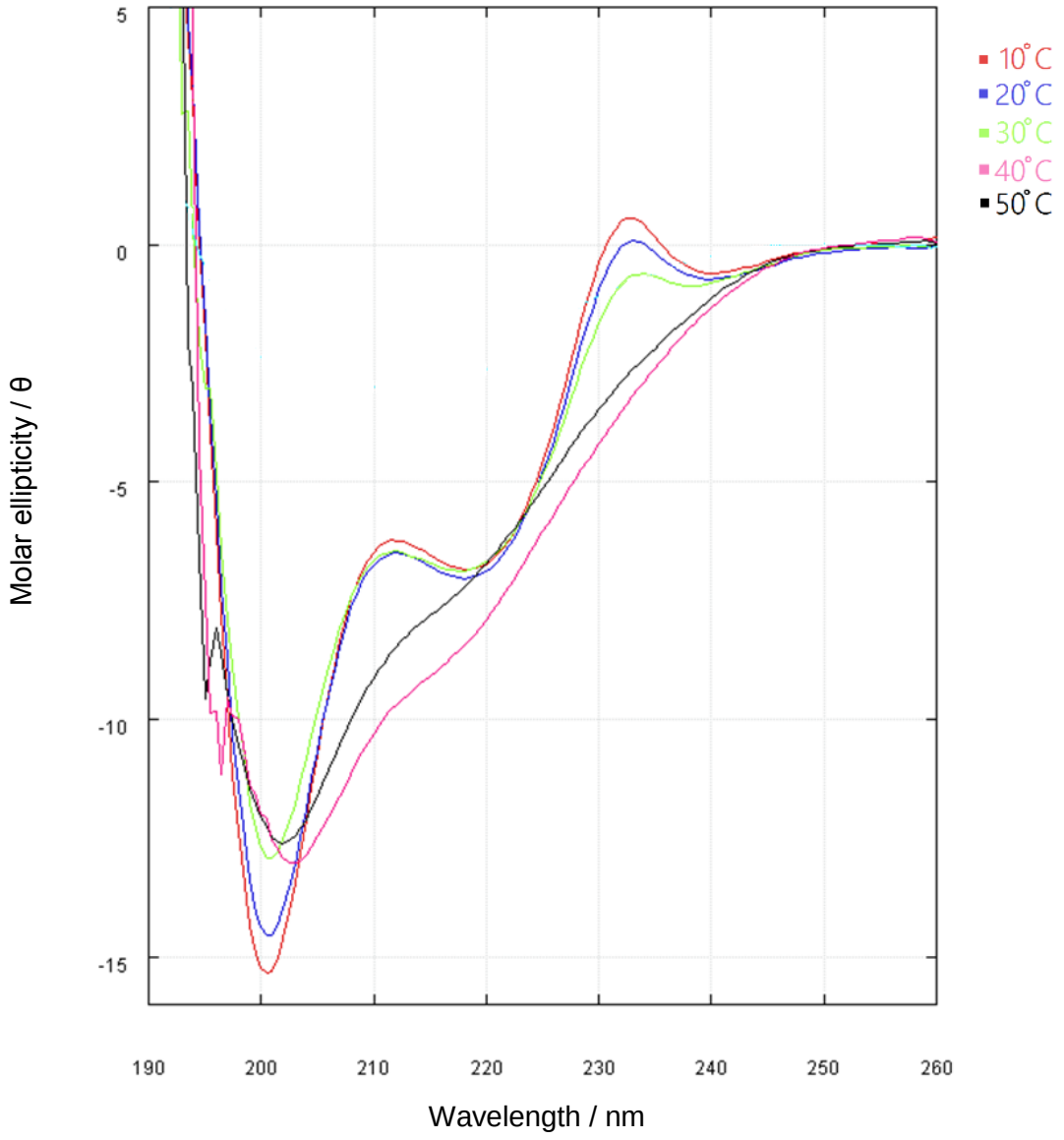


Figure 4.5 CD Thermal Melt Analysis of OutB 112-220
Circular dichroism spectra of purified OutB (residues 112-220) recorded at 10°C (red), 20°C (blue) 30°C (green), 40°C (magenta) and 40°C (black). Spectra are of the same sample in 20mM tris pH7.0. Note the disappearance of the negative peaks at 200nm and 220nm (and a smaller positive peak at 235nm) as temperature increases.

4.5 Dynamic Light Scattering Measurements

Dynamic light scattering (DLS) was used to assess the dispersity of particles within purified OutB protein and suitability for crystallography studies. DLS analysis showed purified OutB 112-220 behaved as a 12kDa protein with polydispersity of 13.9% this indicates that OutB produced by this method is most likely folded and behaves as a monomer in solution (Figure 4.6).

It is generally accepted that proteins with polydispersity of less than 20%, have a very good chance of crystallizing. With 13.9% polydispersity, OutB is a good candidate for crystallization trials. Crystallization trials of the entire periplasmic domain (residues 112-220) did not yield any crystals.

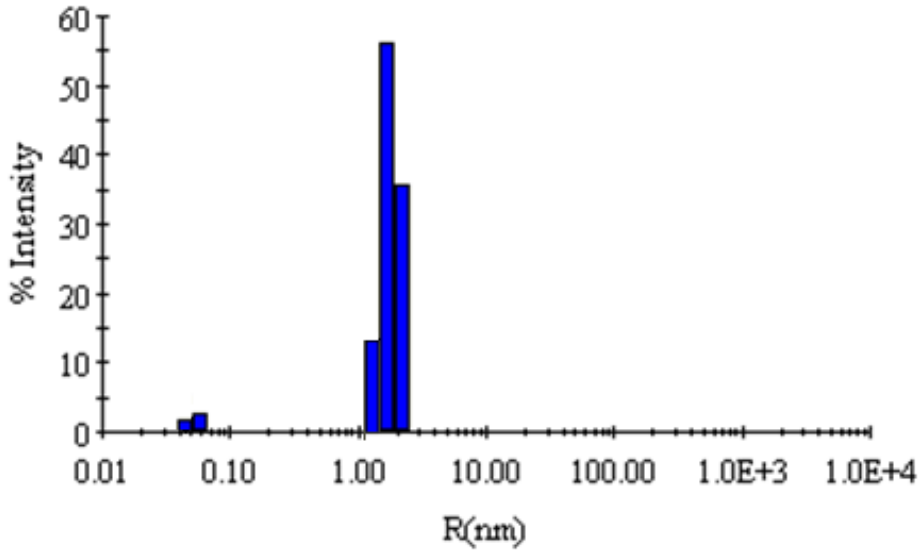


Figure 4.6 Dynamic light scattering of OutB 112-220 reveals low poly-dispersity, compatible with crystallization, but we failed to grow crystals. The hydrodynamic radius is consistent with a monomer in solution.

4.6 Truncation Mutants

Crystallisation trials of the full length periplasmic domain (residues 112-220) yielded no crystals. It was hypothesised that the C-terminal 15 residues might hamper crystallization, perhaps due to flexibility, and were unstructured in solution, and therefore inhibiting crystal formation. To overcome this, truncation mutants were generated by mutating glycine residue codons 193 and 202 (Figure 4.7) (GGA) to STOP codons (TGA) using Quick Change II Mutagenesis.

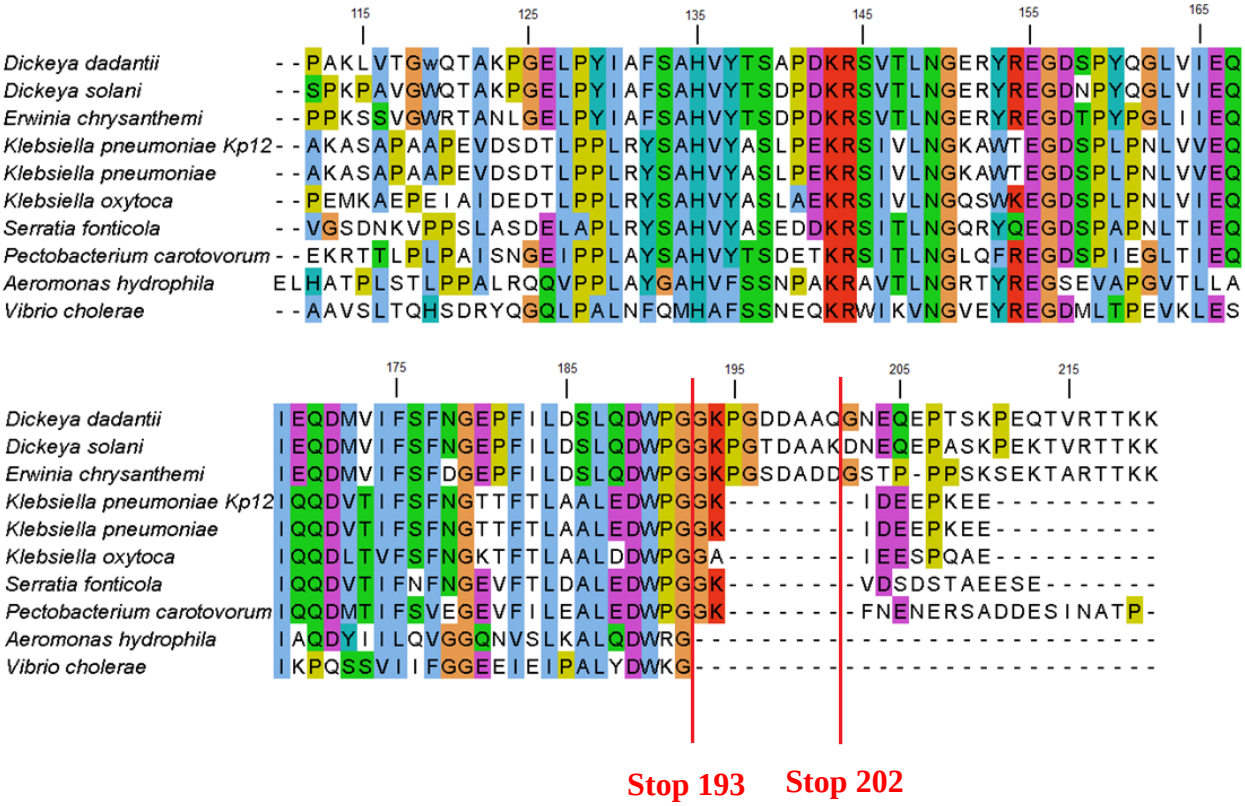


Figure 4.7 Sequence alignment of GspB proteins showing the positions where 'early stop' point mutations were inserted.

An initial crystal hit was obtained using the OutB construct consisting of residues 112 to 202, and screened in 96 well plates, in 400nL drops (200nL 1:1 of protein and precipitant). These were set up by a Mosquito liquid handling robot and obtained by hanging nanodrop vapour diffusion with Molecular Dimensions Structure Screen 1, condition C9 (0.2M $MgCl_2$, 0.1M Tris, 30% w/v PEG4000, pH8.5).

Optimisation of this condition was carried out to determine if crystals could be more readily obtained in a similar condition. This yielded similar crystals through the pH range 8.0 – 8.5 and no crystals at PEG4000 concentrations below 20%. It was therefore decided to scale up the crystallisation to 1µL drops to aid crystal handling, using the Molecular Dimensions stock solution as crystallisation reagent. However, scaling up, to 1-2µL drops in 24 well plates to make crystal handling easier proved unsuccessful by vapour diffusion alone. Further crystals were however, successfully obtained by whisker streak seeding by using the initial hit as a source of seeds.

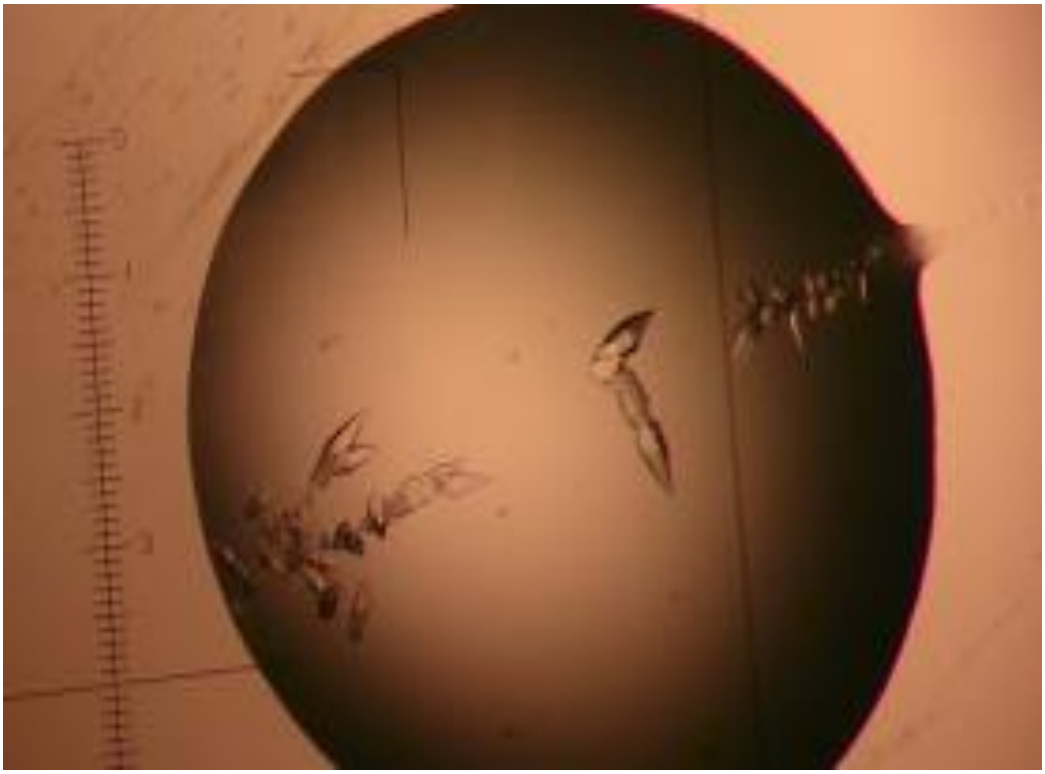


Figure 4.8 OutB 112-202 Crystals
Streak seeding resulted in the crystals used for data collection which were grown from 0.2M $MgCl_2$, 0.1M Tris, 30% w/v PEG4000 at pH8.5. The crystals are approximately 0.07 mm in length**.**

4.7 Data collection

Selected crystals were removed from the mother liquor and transferred to a solution containing mother liquor plus 5% propane-1,2-diol as a cryoprotectant and flash cooled in liquid nitrogen. Crystals were transferred and mounted using 0.1mm litholoops. Diffraction data were collected at 100K using a Pilatus 300M detector at beam line I04 at Diamond Light Source (DLS), Oxfordshire. Two diffraction images were initially recorded at 0° and 90° to assess the quality of the crystal and determine a suitable data collection strategy.

Data were integrated and scaled by XDS through the Xia2 (Winter, 2009) software package available through DLS. Several data sets were collected from different crystals (Appendix C) and assessed for quality based on statistics such as maximum resolution, I/σ, R_{merge} and R_{pim}. All data sets were highly similar, with only small differences in the calculated statistics, as provided by XDS. The best data set was chosen for molecular replacement. Statistics given in Table 4.1. A total of 1200 images were collected over a rotation range of 120° with a 0.1° oscillation angle per image.

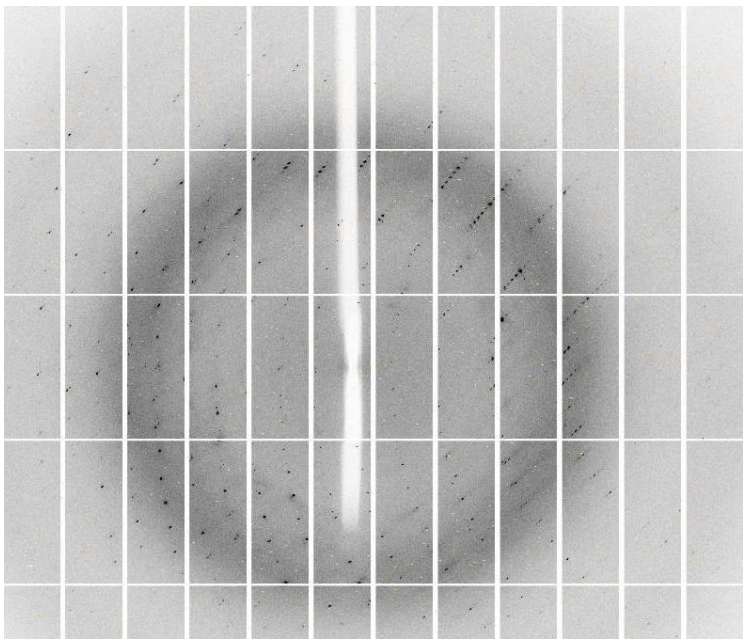


Figure 4.9 Diffraction image obtained from an OutB crystal
A sample diffraction image obtained from an OutB crystal showing clearly defined spots.

Table 4.1 Crystallographic data statistics for OutB crystal used for model building and refinement

Data collection	
Space group	P 4_3 2_1 2
Cell parameters *a*, *b*, *c* (Å)	*a* = *b* = 36.3, *c* = 133.2
α, β, γ (°)	$\alpha = \beta = \gamma = 90$
Molecules per asymmetric unit	1
Wavelength (Å)	1.07
Resolution (Å)	2.02-36.30 (2.02-2.08)
Total number of observations	56965 (2403)
Number of unique reflections	6380 (435)
Multiplicity	8.9 (5.5)
Completeness (%)	99.2 (93.6)
R_{merge} (%)	3.6 (57.4)
Mean <I/σ(I)>	30.6 (3.7)
Wilson B-factor (Å^2)	44.4
R_{pim}(I)	0.014 (0.28)
R_{meas}	0.041 (0.69)

Values for the highest resolution shell are given in parentheses.

4.8 Molecular replacement

Ab initio modelling of the OutB sequence by the PHYRE2 server (Kelly, 2009) suggested that the structure of OutB may be similar to the structure of the homology region (HR) domain of GspC. On this basis, and given that structures of GspC-HR domains are available, molecular replacement was tested to see if a solution could be achieved. Trial and error manipulation of the enterotoxigenic *E. coli* GspC-HR structure (PDB code: 3OSS), including truncation and modelling the protein as poly-serine, eventually resulted in a PHENIX Phaser (McCoy, 2007) TFZ score of 7.9, approaching the value of 8.0 or greater expected for a genuine solution.

4.9 Model building and refinement

PHENIX Autobuild (Terwilliger, 2007) built 7 chains comprising a total of 59 residues of the sequence of GspB$^{112-202}$ into the electron density map giving R-factor and R-free values of 0.42/0.50, respectively. Twenty cycles of manual rebuilding and refinement of the structure of *D. dadantii* GspB$^{112-202}$ yielded a model with reasonable stereochemistry, given the high proportion of irregular structure. Of the 90 residues of OutB$^{112-202}$ crystalised, residues 115-198 could be built into the electron density (Figure 3.9); the first three and the last four residues are probably disordered or have multiple conformations in the crystal and are not seen.

Initial assessment of the data by XDS suggested space group $P4_12_12$, however, reprocessing of the data in space group $P4_32_12$ resulted in lower R-factors and is therefore the more likely space group.

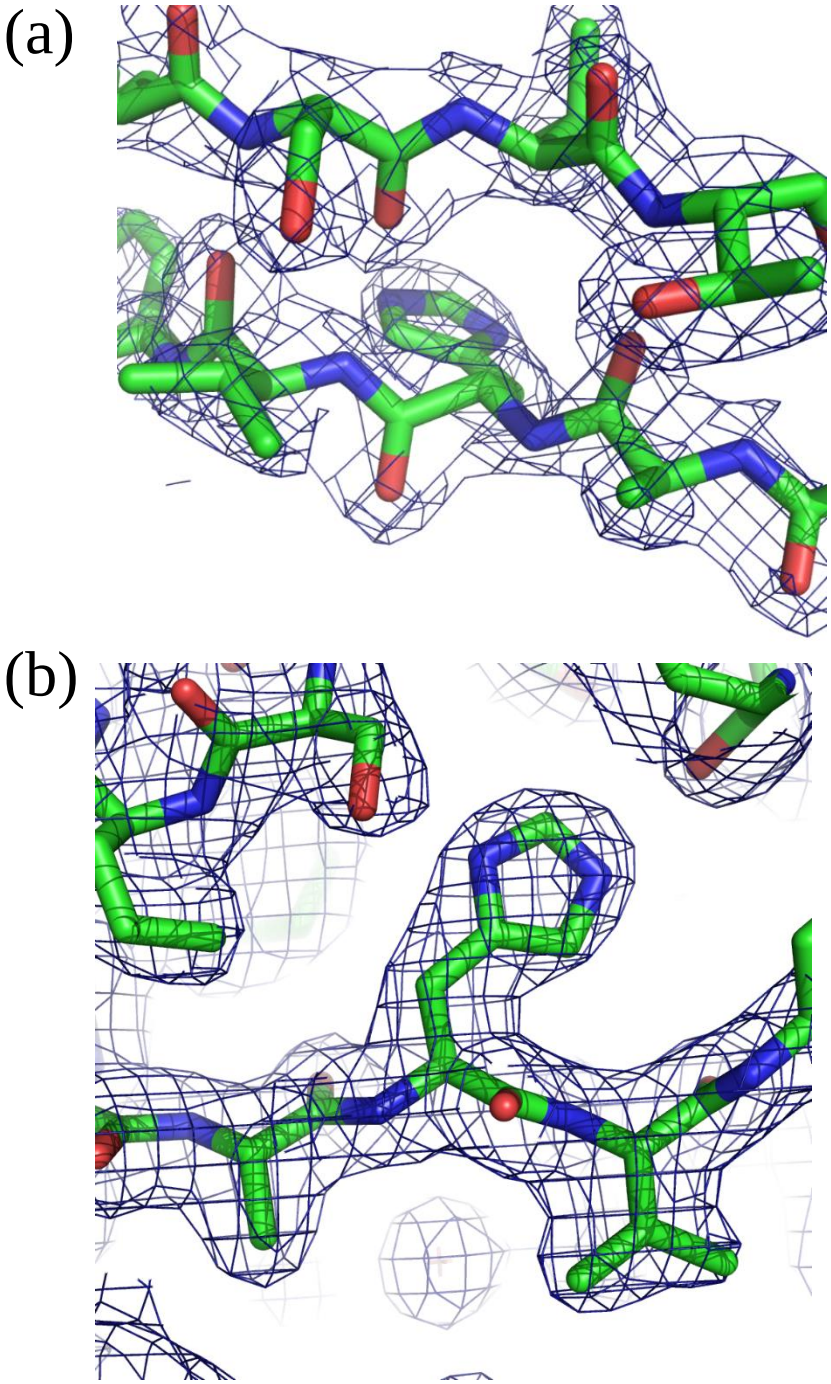


Figure 4.10

(a) σ_A-weighted $2F_{obs}$-F_{calc} Fourier synthesis, contoured at 1.2σ around β-strands 1 and 2. The lower strand is β1 running right to left (residues: A134, H135, V136, Y137) and the upper one is β2 running left to right (residues: S145, V146, T147).

(b) Same Fourier synthesis map and residues, showing H135 at the centre V136 lower right and the electron density around the indole ring and alkyl side chain of the respective residues

4.10 Structure Validation

The stereochemistry of the structure was checked by MolProbity and PROCHECK (Laskowski et al. 1993) by plotting the Ramachandran plot. Bond lengths and bond angles of main chain and side chains, planarity, chirality, solvent content and torsion angles of main chain (phi and psi) were monitored by this program. Any distorted geometry was corrected manually using COOT.

The root-mean square deviations (rmsd) from ideal values of bond lengths and bond angles are 0.013Å and 1.68° respectively with all residues having acceptable phi, psi values (Table 1). The final refinement of the structure yielded R-factor/R-free of 0.209 and 0.254 respectively at 2.05 Å resolution (Table 4.2).

Table 4.2 Refinement statistics for OutB

Refinement	
Resolution limits (Å)	2.05-34.91
R-factor (%)	20.85
R-free[b] (%)	25.42
RMSD bonds (Å)	0.015
RMSD angle (°)	1.81
Average B-factor (Å^2)	33.38
Number of protein atoms	666
Number of solvent atoms	15
Ramachandran plot statistics	
Residues in most favoured regions (%)	100

4.11 Description of the architecture and comparison with GspC-HR and PilP domains

The polypeptide chain of the periplasmic domain of GspB forms two three-stranded antiparallel β-sheets; the β-strands sequentially form the up-down-up β-sheets, so the first sheet comprises strands 1, 2 and 3 and the second 4, 5 and 6 (Fig. 3.10). The two sheets are at approximately 70° to each other so the structure forms a small barrel with a hydrophobic core. There are three residues in 310-helix conformation in the loop between β1 and β2.

Perhaps the most striking feature of the structure is the quantity of irregular polypeptide, between the short β-strands, especially but not limited to the β3/ β4 loop, and at the amino- and carboxy-ends of the small β-barrel (Figure 4.11). DSSP (Kabasch 1983) reveals 58% of the residues do not fall into a recognized secondary structure category while 38% of residues form β-sheet and 4% form 310-helix. The sequential β-strands are: β1 Ser133-Val136, β2 Ser145-Leu148, β3 Glu151-Tyr153, β4 Leu163-Gln167, β5 Met172-Ser176 and β6 Glu180 –Leu184. Pro141 to Lys143 form the 310-helix. The 310-helix in the loop between β1 and β2 contains a highly conserved Lysine 143 and Arginine 144 which forms a salt bridge to Glu155, also highly conserved. Lysine 143 does not form any contacts within the molecule and therefore is likely to have a functional role.

A search for structural homologs using the DALI webserver (Krissinel, 2004) returned: the enterotoxigenic E. coli GspC-HR domain used as the molecular replacement search model (PDB code:3OSS, Z-score 7.8, rmsd 1.7Å for 60 equivalent CA atoms, sequence identity 17%); *Pseudomonas aeruginosa* PilP from the type 4 pilus system (PDB code:2Y4Y, Z-score 5.6, rmsd 2.8Å for 60 equivalent CA atoms, sequence identity 12%); and Dickeya dadantii OutC-HR (PDB code :2LNV, Z-score 3.5, rmsd 2.8Å for 58 equivalent CA atoms, sequence identity 10%) as related structures (Figure 3.11).

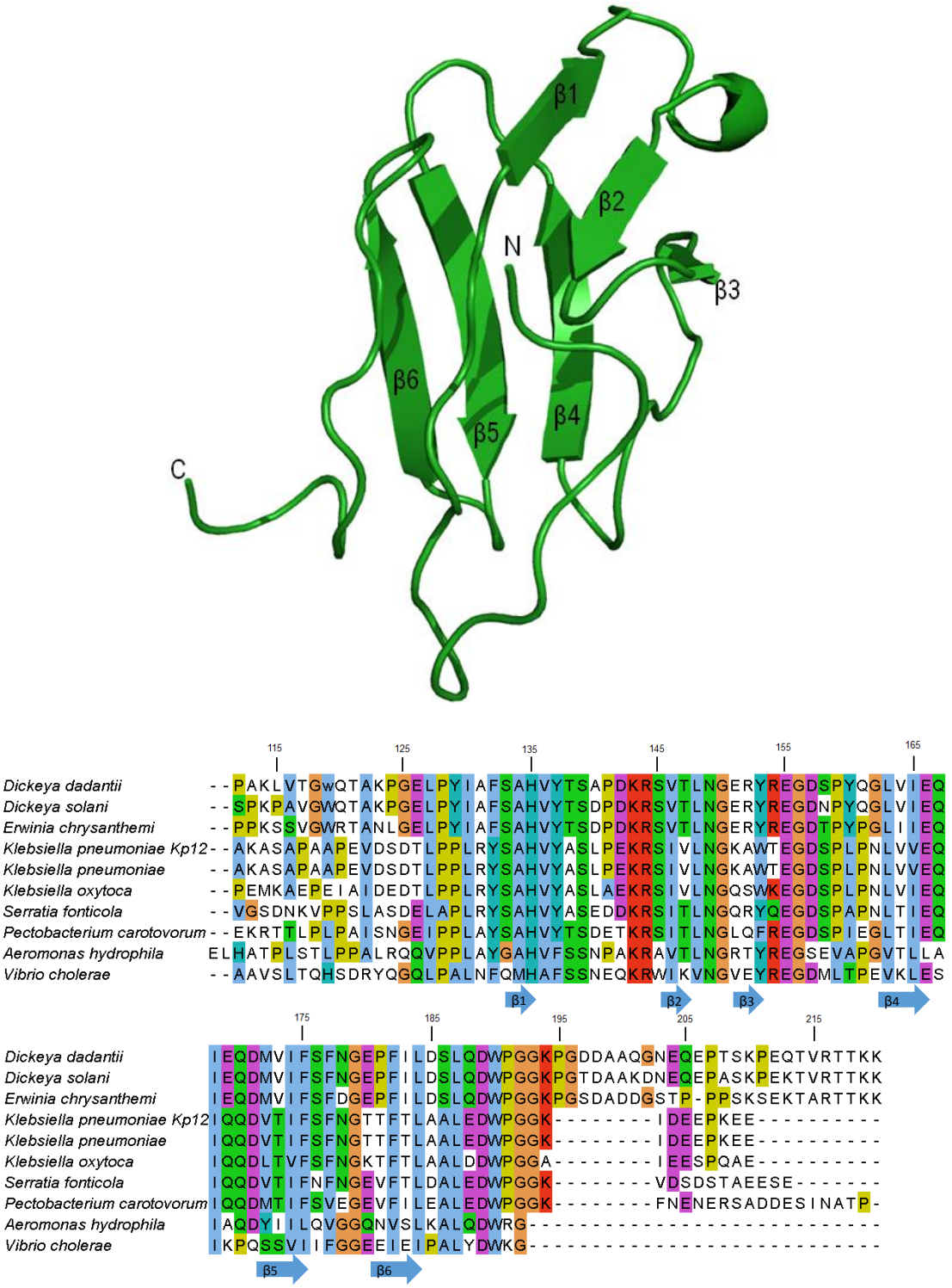


Figure 4.11 Cartoon structure of OutB

Top: A cartoon of the final OutB structure shown in ribbon form with the β-strands labelled. Bottom: A sequence alignment of various GspB proteins, with the position of the β-strands present in the model annotated below the sequence as arrows.

It is interesting that *D. dadantii* GspB is more similar to enterotoxigenic E. coli GspC-HR in structure than to *D. dadantii* GspC-HR, suggesting that D. dadantii GspB and E. coli GspC-HR may be closest in function.

The results show that the structure of GspB and *E. coli* GspC-HR are closely similar supporting the view that the role of *D. dadantii* GspC-HR in assembling the secretion system has been assumed in part or whole by GspB, possibly allowing *D. dadantii* GspC-HR to become more specialised in substrate selection.

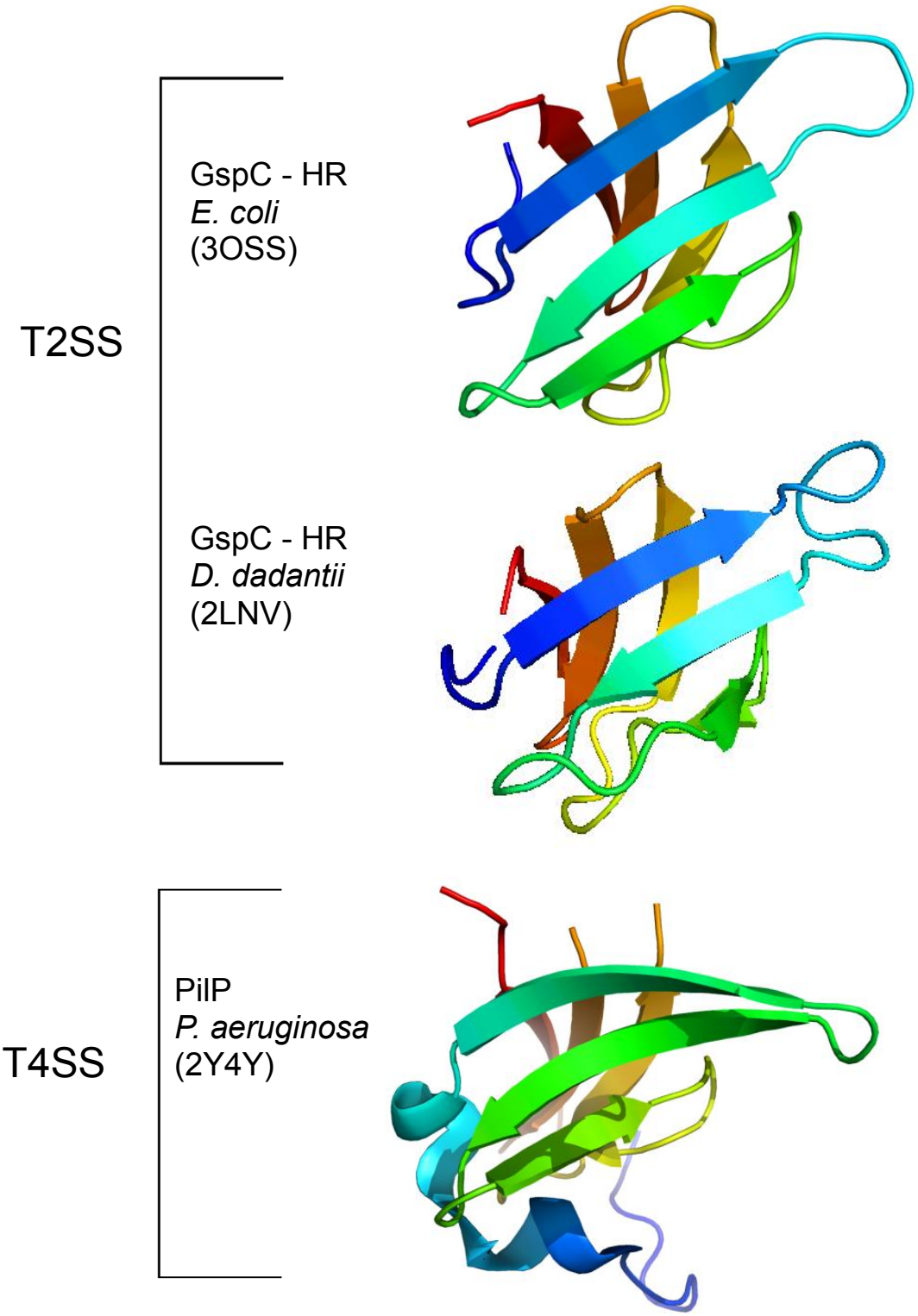


Figure 4.12 Structural homologues of *D. dadantii* OutB. (a) *E. coli* GspC-HR (PDB code: 3OSS) (b) *D. dadantii* OutC-HR (PDB code: 2LNV) and (c) *P. aeruginosa* PilP (PDB code: 2Y4Y). The cartoons are coloured from N-terminal end (blue) to the C-terminal end (red) of the polypeptide chain and the proteins shown in a similar orientation to OutB in Figure 4.11 (*i.e.* with the surface formed by β strands 1-3 at the front)

4.12 Role of GspB – Structure Function Analysis

Enterotoxigenic *E. coli* GspC-HR has a strong interaction with the N0 domain of the secretin, GspD (Korotkov, 2011) in comparison to that observed with the corresponding proteins in *D. dadantii* (Gu et al., 2012a). It is possible that the role of assembling the functional type II secretion system fulfilled by GspC-HR in the enterotoxigenic *E. coli* has been partially taken by GspB in *D. dadantii* freeing up D. dadantii GspC-HR for a greater involvement in other functions, such as determining the specificity of substrates secreted by the system.

Calculating the electrostatic surface potentials of OutB reveals a central band of positive potential with negative bands either side (Fig. 4.13a). Its structural homologues, *D. dadantii* and E. coli GspC-HR domains, have a neutral central band with negative potential either side and are quite similar to one another (Fig. 4.13b, c), while the periplasmic domain of *P. aeruginosa* PilP is more similar to GspB in terms of electrostatic potential, having a pronounced central band of positive potential (Fig. 4.13d). The striking similarity of PilP to GspB both structurally and electrostatically suggests PilP they have similar functions in the type II secretion and type IV pili systems respectively.

Analysis of the crystallographic symmetry reveals no clues as to the *in vivo* organisation of OutB. The protein is stable as a monomer *in vitro* as shown by the nature of the crystal and through size exclusion measurements (Appendix D). This is unusual for such a protein which forms part of a complex, multiprotein system in which interactions are numerous.

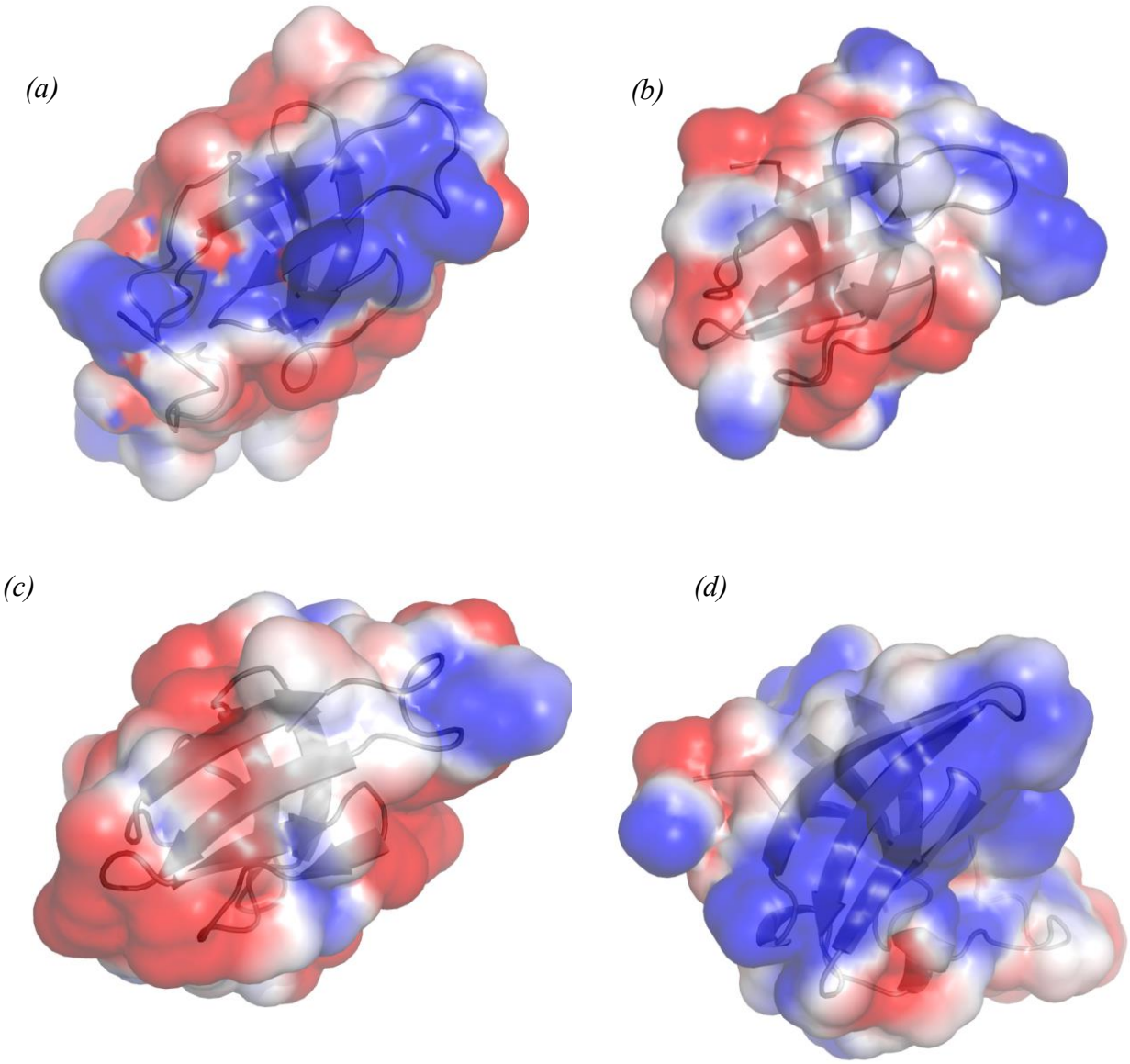


Figure 4.13 The electrostatic potential, calculated using macroscopic dielectric constants and a finite-difference solution of the Poisson-Boltzman equation calculated using APBS software (Dolinsky, 2007). Potentials are mapped to the surfaces of (a) *D. dadantii* GspB (b) *E. coli* GspC-HR (PDB code: 3OSS) (c) *D. dadantii* GspC-HR (PDB code: 2LNV) and (d) *Pseudomonas aeruginosa* PilP (PDB code: 2Y4Y). The potential surface is contoured at +1kT/e (blue) and -1kT/e (red).

Chapter 5: Interaction studies of OutB

5. 1 Aim

Pull down assays show a reasonably strong interaction between OutB and the N0 domain of OutD (Shevchik *et al.* unpublished – Figures 5.1 & 5.2). Experimental studies on the GspC proteins which are known to interact with this domain of the GspD secretin (OutC-HR with OutD-N0 in *D.* dadantii and GspC-HR and GspD-N0 in enterotoxigenic *E.* coli) suggest two possible modes of interaction: β1-β1 in the crystal structure of Hol *et al.* 2011 or β1-β3 in the NMR experiments of Gu *et al.* 2012. Here, I introduce cysteine residues on the β 1 strand of N0 and β 1 and β 3 strand of OutB in an attempt to trap the complex formed by OutB and OutD-N0 and reveal the mode of interaction in vitro.

5. 2 Background to OutB / OutD interaction

Work by Shevchik in 2001, indicated that OutB is associated with the outer membrane through its C-terminal portion. This suggested an interaction with a protein of the outer membrane, as the OutB protein is integral to the inner membrane. Unpublished pull-down assays by Shevchik *et al* (Figure 5.2) confirmed the domains responsible for this interaction (Figures 5.1 & 5.2).

The C-terminal domain of OutB (residues 112-220) co-purifies with the N-terminal domain N0 (residues 28-112). Given the structural similarity of OutB to the HR domain of GspC (Chapter 4) which also interact with the same domain of their respective GspD secretins, it was hypothesized that OutB may interact with OutD-N0 domain in a similar way. However, this is complicated by the fact that 2 different interactions between GspC-HR and GspD-N0 domains have been observed.

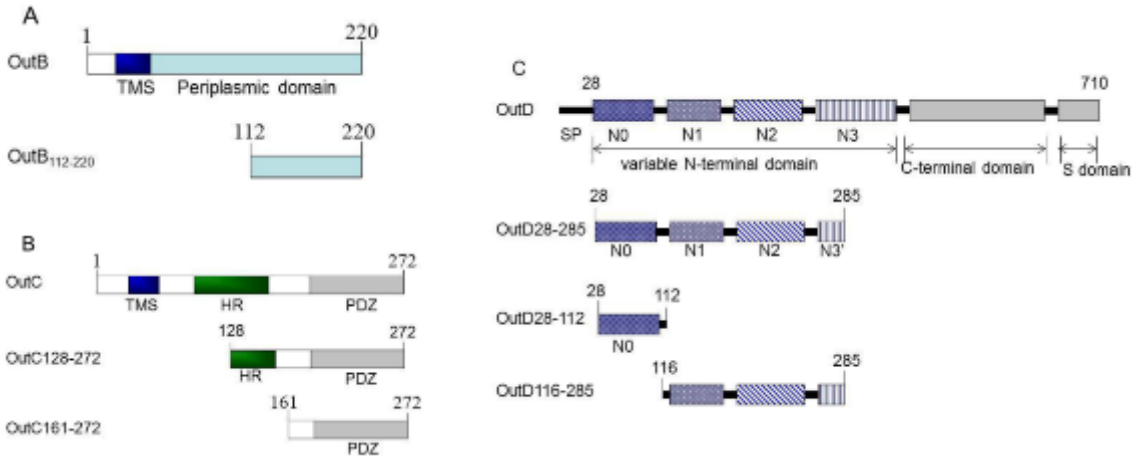


Figure 5.1 Schematic diagrams of OutB (A), OutC (B) and OutD (C) and their truncated derivatives used in the pull down assays.

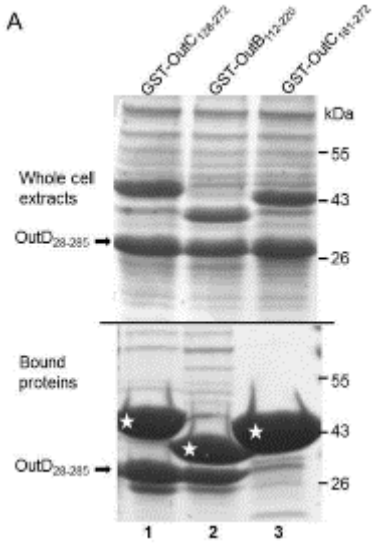


Figure 5.2 Mapping the interaction regions between OutB and OutD.

Upper panel, whole-cell extracts of E. coli BL21 (DE3) co-expressing GST-OutB112-220 or GST-OutC derivatives (upper protein bands, corresponding names are indicated on top) with OutD derivatives (corresponding protein bands are indicated with an arrow).

Lower panels, the extracts from the upper panels were used in pull-down assays on Glutathione Sepharose. Bound proteins were eluted with sample buffer, separated by SDS-PAGE and stained with Coomassie. GST-OutB112-220 or GST-OutC derivatives are indicated by asterisks, OutD derivatives, by arrows.
GST-OutC128-272 and GST-OutC161-272 were used as a positive and a negative controls for OutD binding, respectively. Note that OutD28-116 and OutD28-285 but not OutD116-285 were bound to GST-OutB112-220.

5. 3 Models of the OutB / N0 interaction based on GspC-HR / GspD-N0

5.3.1 β1 – β1 Hol *et al.* 2011

Hol *et al.* 2011 have published the HR domain of GspC co-crystalised with the GspD-N0 domain from *Escherichia coli* (PDB: 3OSS). In this interaction (Fig. 5.3), the interface between GspC-HR and GspD-N0 buries 1280 $Å^2$ of accessible surface area with a calculated ΔG of interaction of -5.4 kcal×mol^{-1} as assessed by the PISA server (Krissinel and Henrick, 2007).

The overall shape of the interface is relatively flat with a small concave area on the GspD surface. A total of 18 residues from GspC-HR and 19 residues from GspD-N0 engage in a combination of hydrophobic interactions and hydrogen bonds. The first three β strands of GspC-HR and the first β strand plus the subsequent α helix (α1) of GspDN0 are the major contributors to the interface. The majority of the hydrogen bonds are formed by an antiparallel arrangement of strand β1 of GspC-HR and strand β1 of GspDN0 (Table 4.1). This β-strand augmentation is frequently observed in protein–protein interfaces (Remaut and Waksman, 2006). The residues involved in forming β-strand interactions are listed in Table 4.2.1.

GspD-N0	OutD-N0 equivalent	GspC-HR	OutB equivalent
T 6 A 7 N 8 F 9 K 10	F 3 S 4 A 5 S 6	F134 A 133 I 132 G 131 R 130	H 135 V 136 Y 137 T 138

Table 5.1 Residues involved in the interaction published by Hol et al. 2011 and the equivalent residues that would be involved if the interaction between OutD and OutB was identical.

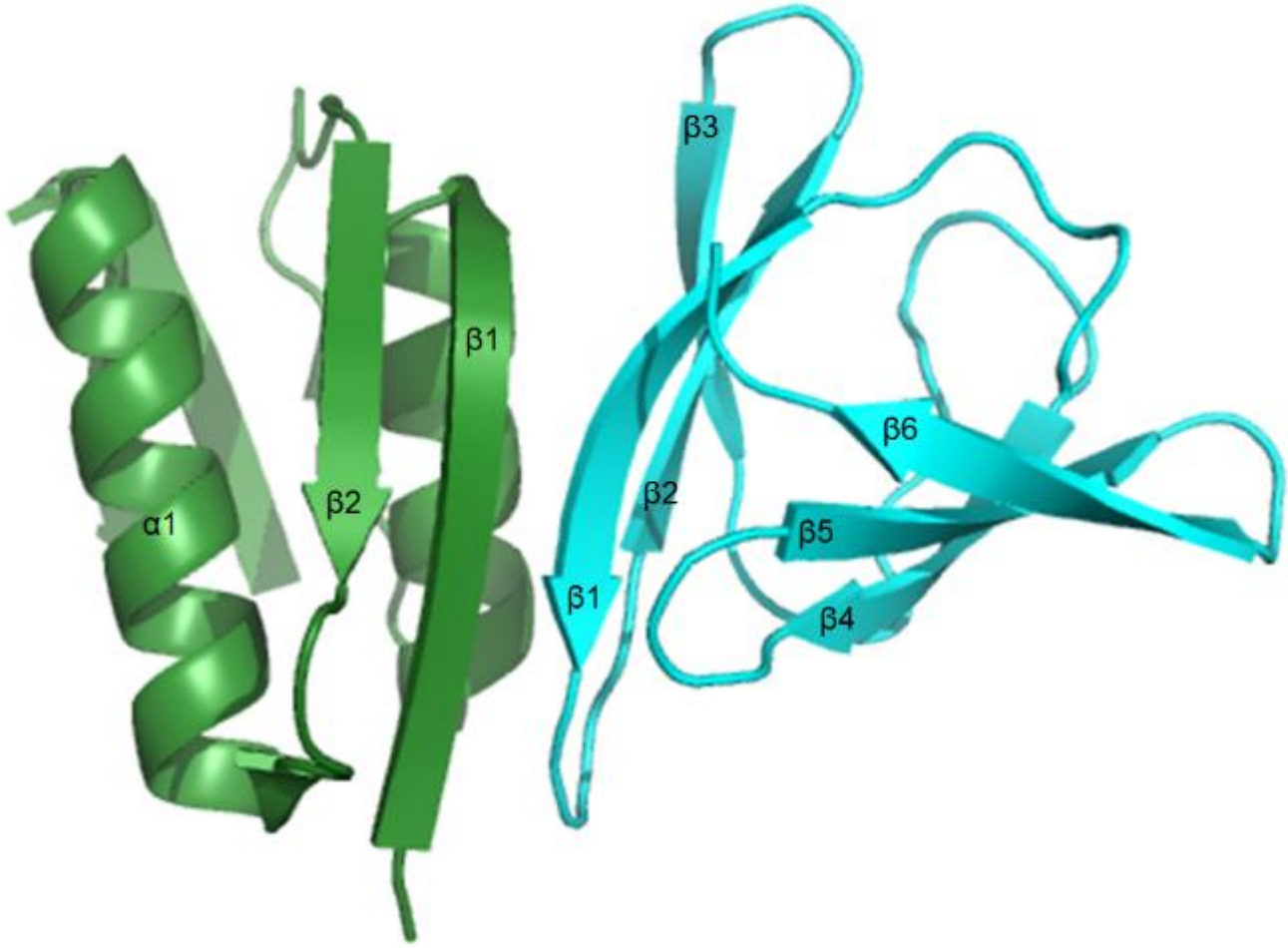


Figure 5.3 A cartoon representation of GspC-HR and GspD-N0 from enterotoxigenic *E. coli* as observed in the crystal structure of Hol *et al.* 2011 (PDB: 3OSS) Note the antiparallel arrangement of the best strands β1 from each protein at the interaction interface.

5.3.2 β1 – β3 Gu *et al.* 2012

An alternative mode of interaction of the GspC-HR and GspD-N0 domains in presented by Gu *et al.* 2012a (Fig. 5.4). In their experiments the homologous proteins from the *D. dadantii* Out system were used (OutC-HR and OutD-N0). The model was proposed on the basis of NMR chemical shift mapping. In the experiments, one protein was isotopically labelled, the other NMR silent. The chemical shifts observed when unlabeled HR domain was added were mapped to the surface of the N0 domain and *vice versa* to determine the 2 surfaces involved in the interaction.

The chemical shifts observed reveal that residues on strands β1, β2, and β3 of the HR domain are involved in the interaction and residues on β1 of N0 are most affected (Table 4.2). Residues on β1 and α2 of the N0 domain were the most affected, but the shifts were relatively small. Small shifts were seen for residues: Thr63, Ile64, Ser65, Phe79, Ser83, and Val84.

The model brings β1 of HR and β3 of N0 together such that they could form a continuous anti-parallel β-sheet across both domains HR and N0. It is important to note that this is only a model of the interaction and the position and conformation of interfacial residues has not been defined experimentally. Met101 appears to be an important hydrophobic residue at the interface with the more remote Asp141 also showing large chemical shifts. The interaction between the *D. dadantii* HR and N0 domains is weak, but it does persist in the presence of 150 mM sodium chloride.

OutD-N0	OutC-HR	OutB equivalent
Thr 63	Thr 98 - β1	A 134
Ile 64	Gly 99 - β1	H 135
Ser 65	Val 100 - β1	V 136
Phe 79	Met 101 - β1	Y 137
Ser 83	Ala 102 - β1	T 138

Table 5.2 Residues involved in the interaction published by Gu *et al.* 2012a and the equivalent residues that would be involved if the interaction between OutD and OutB was identical.

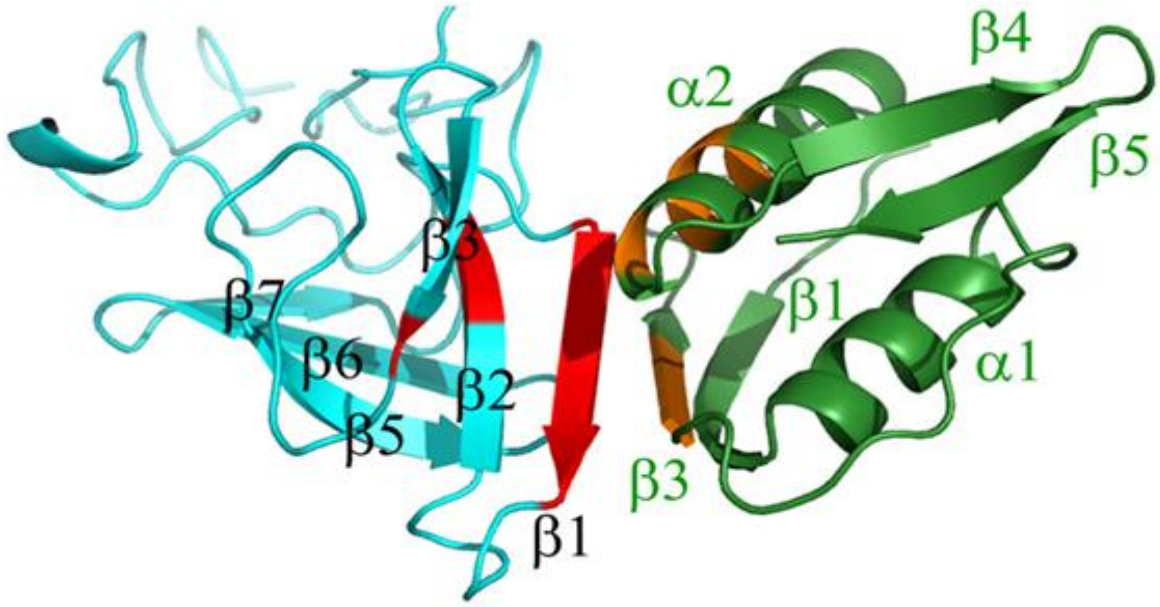


Figure 5.4 A cartoon representation of GspC^{OutC}-HR and GspD^{OutD}-N0 from *D. dadantii* as proposed on the basis of the NMR experiments of Gu *et al.* 2012a (GspC^{OutC}-HR structure PDB: 2LNV / GspD^{OutD}-N0 homology model from Modeller v9.1) Note the antiparallel arrangement of the best strands β1 from each protein at the interaction interface.

N.B. In several members of the HR domain family of structures, there is a fourth β-strand in the second β-sheet, hence the numbering scheme misses β4. This strand is more dynamic in solution than the other six β-strands, which are present in all of the NMR ensemble structures

5.4 Generation of cysteine mutants of OutB and N0

The structure of GspB as determined in Chapter 4, was superimposed over the GspC-HR domain in both of the previous structures, and the equivalent residues from β1most likely to be involved in an interaction were determined to be valine 136, tyrosine 137 and threonine 138. In the construct encoding only the periplasmic domain of GspB, residues 112-220, these were V25, Y26 and T27 respectively.

Using the same process, the corresponding residues of OutD on either β1 (serine 4, alanine 5 and serine 6) and β3 (threonine 63, isoleucine 64 and serine 65) were selected and made using Strategene Lightnigh Quick Change II method (Chapter 3).

5.5 Disulphide cross-linking studies with addition of crosslinking reagents

Homobifunctional, maleimide compounds (Figure 5.5) can be used as crosslinkers for conjugation between sulfhydryl groups (-SH) which are present in the R-groups of cysteine

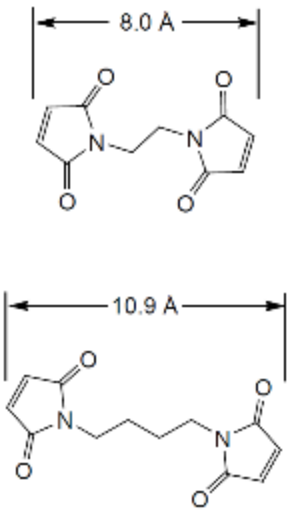


Figure 5.5 Maleimides
Top: Bismaleimidoethane (BMOE) is capable of crosslinking sulphydryl groups 8Å apart.
Bottom: Bismaleimidobutane (BMB) can span 10.9Å

residues. Such bismaleimide crosslinkers are commonly used to explore and characterize protein structure (i.e., oligomerization) or protein interactions.

Because BMOE and BMB have the same reactivity but differ in length, the relative success of these three reagents in forming crosslinks between sites in a protein oligomer or interaction can assist in determining intra- and intermolecular distances.

Reaction of a sulfhydryl to the maleimide group results in formation of a stable thioether linkage (Figure 5.5), which cannot be cleaved by reducing agents or physiological buffer conditions. Reaction of maleimides is very specific to sulfhydryls at pH 6.5 - 7.5. Although maleimides will react to primary amines at pH > 8, the rate is 1000 times slower than the reaction to sulfhydryls at pH 7. Unlike iodoacetamides, maleimides do not react with tyrosines, histidines or methionines, so it is possible to ensure crosslinking only occurs between surface exposed cysteine residues.

5.6 Results

5.6.1 OutD-N0 homodimers – A useful positive control

Each of the mutants detailed in Table 5.3 below were purified separately to individually optimize expression of each mutant and to confirm the solubility of the purified protein. Subsequently, controls were run of the cysteine mutant proteins under both reducing and oxidizing conditions.

OutD-N0 Mutants		**OutB Mutants**	
S4 A5 S6	β1	V136C Y137 T138	β1
T63 I64 S65	β3		

Table 5.3 A list of the mutants generated A list of the mutants generated to test interactions between β1 of OutB and both β1 and β3 of OutD-N0

Initially cysteine mutant proteins were purified in the same buffer as the non-mutant form. However, it was discovered that mutations S4C, A5C and S6C of OutD-N0 formed homodimers during the purification process, so it was necessary to add DTT to the lysis and size exclusion buffers.

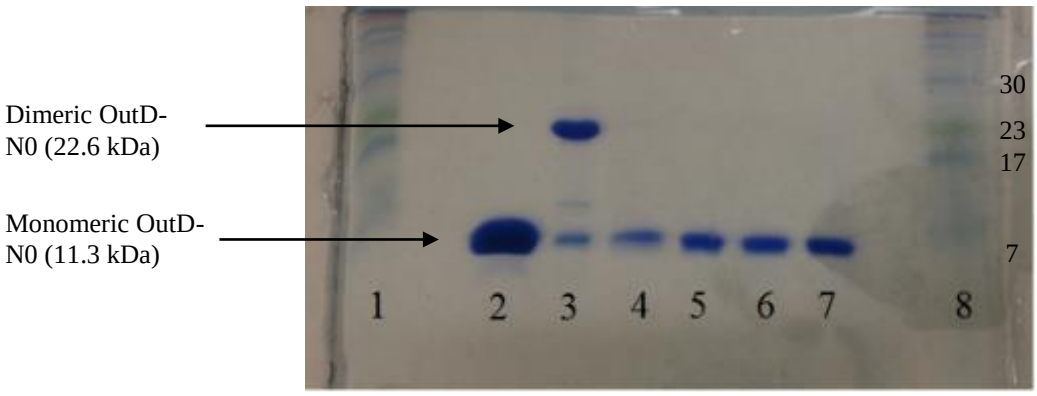


Figure 5.6 20% SDS-PAGE gel showing the addition of DTT prevents the formation of OutD-N0 mutant S4C from forming homodimers. Marker ladder was loaded into lanes 1 and 8. **2**. non-mutant OutD-N0, **3**. OutD-N0 S4C **4**. 3 + 2mM DTT **5**. 3 + 5mM DTT **6**. 3 + 10mM DTT **7**. 3 + 20mM DTT

5.6.2 Difficulty in distinguishing OutB/OutD heterodimer from OutD homodimers

Due to the similarity in size between the 2 domains (OutD-N0: / $OutB_{112-202}$:) it was difficult to determine by SDS-PAGE analysis alone, if the higher molecular weight species present upon addition of glutathione (used as an oxidizing agent) was due to the formation of homo- or heterodimers. Therefore, mutant OutB proteins were purified leaving the GST tag intact, thereby increasing the mass of the respective OutB domain to 47kDa (36kDa glutathione s-transferase tag + 11kDa for the functional domain).

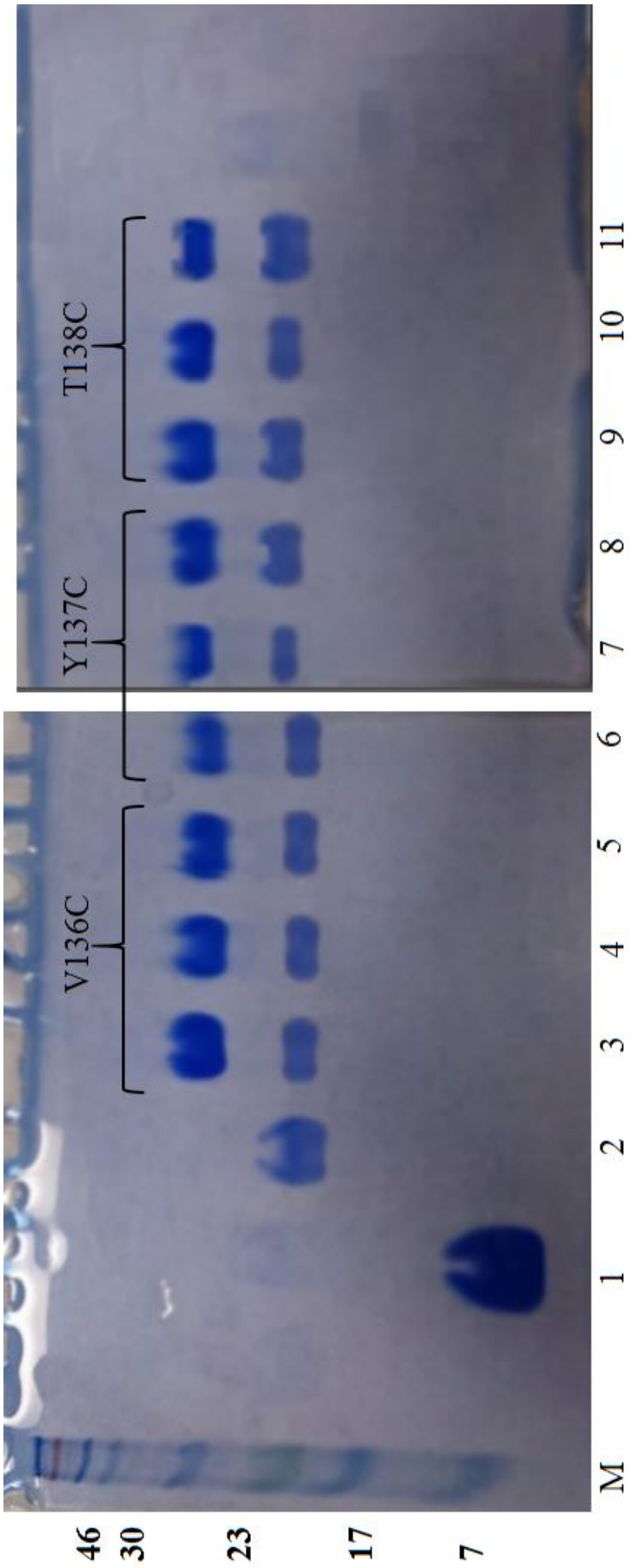


Figure 5.7 Testing the β1 – β1 interaction model for OutB and OutD-N0
Molecular weight markers are shown on the left (lane M; weights in kDa)
Control samples Lane 1. OutD-N0;S4C: under reducing conditions, Lane 2. OutD-N0;S4C: under oxidising conditions.
The following lanes are all under oxidising conditions, containing both OutB (indicated by parentheses above lanes) and the following OutD-N0 cysteine mutants:3. S4C, 4. A5C, 5. S6C; 6. S4C, 7. A5C, 8. S6C; 9. S4C, 10. A5C, 11. S6C

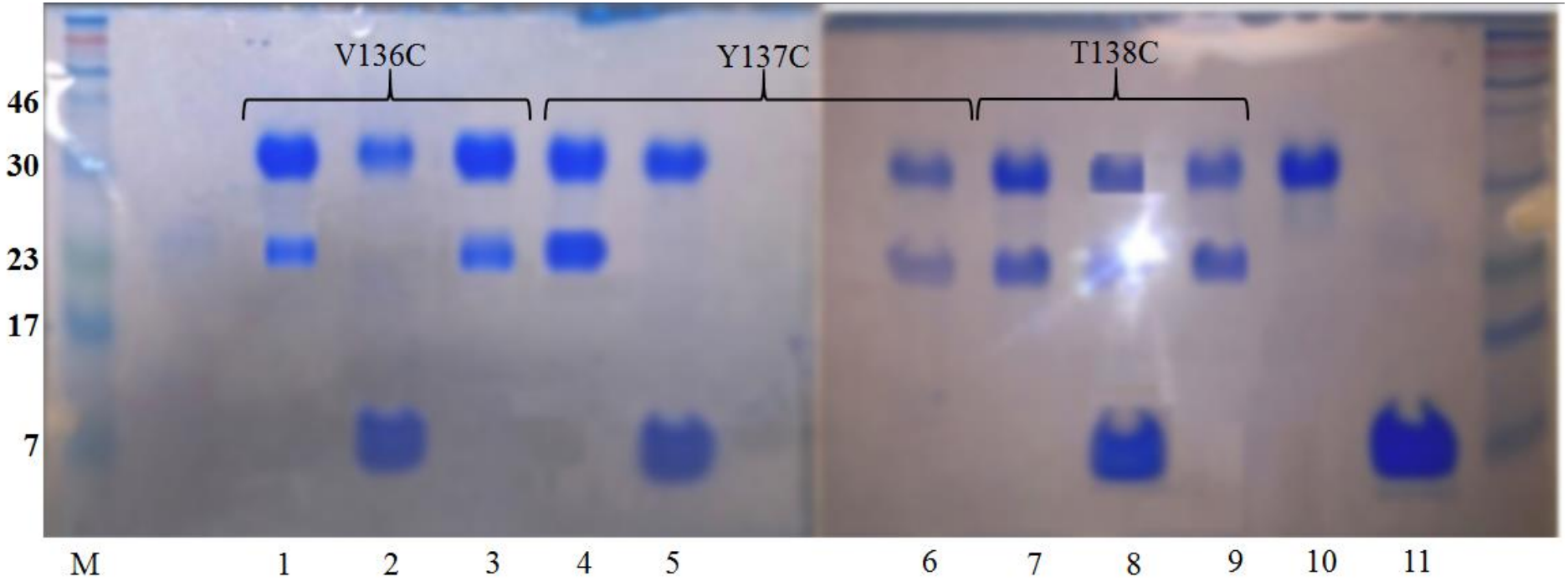


Figure 5.8 Testing the β1 – β3 interaction model for OutB and OutD-N0
Molecular weight markers are shown on the left (lane M; weights in kDa)
Control samples Lane 11. OutD-N0;S4C: under reducing conditions, Lane 10. OutB:V136C: under oxidising conditions.
The following lanes are all under oxidising conditions, containing both OutB (indicated by parentheses above lanes) and the following OutD-N0 cysteine mutants:1. T63C, 2. I64C, 3. S65C; 4. T63C, 5. I64C, 6. S65C; 7. T63C, 8. I64C, 9. S65C

5.6.3 Addition of crosslinking reagents to aid disulphide formation

As shown in figures 5.7 and 5.8, no crosslinking was observed – indicated by the absence of any proteins with a molecular mass greater than GST-OutB monomer (36,482 Da). In order to aid in the formation of covalent bonds between GST-OutB and $OutD^{N0}$, bismalemide crosslinking reagents (BMOE and BMB) were added to attempt to bridge the gap between any cysteines which were close in space, but not quite close enough to form disulphides, also failed to link the 2 proteins together. However, this also failed to link the two proteins.

In order to test the validity of this experiment, and confirm that the cysteine was available to the bismalemide reagent, the mass was determined of protein treated with BMB under, oxidising conditions, using liquid chromatography – mass spectrometry (ESI-MS) (Figures 5.9 and 5.10).

The mass of a single species present in solution was determined to have a molecular weight of 36,739/36,740. This is very close the expected mass of 36,730 for GST-OutB with a covalently attached BMB molecule on the surface (molecular masses of 36482 and 248 respectively). The same was done for the cysteine mutants of OutD-N0, however, this was inconclusive in confirming addition of the bismalemide crosslinkers under oxidising conditions, possibly due to dimer formation (Data in Appendix E).

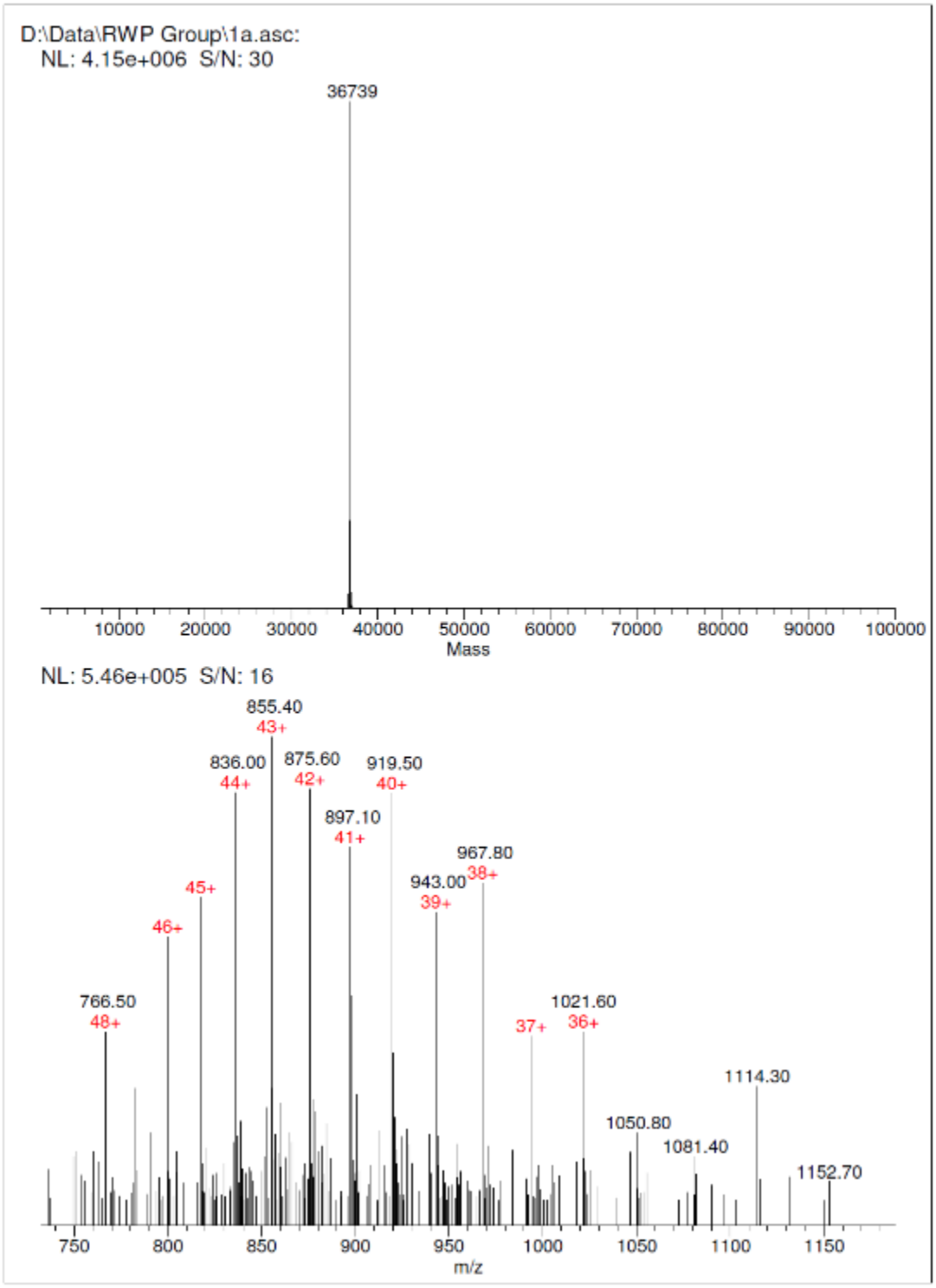


Figure 5.9 Bottom panel: The direct readout from electrospray ionization mass spectrometry of GST-OutB mutant Y137C treated with 1,4-bismaleimidobutane (BMB). The *x*-axis is the mass-to-charge- ratio of the ion, and the *y*-axis is the relative intensity of the species. **Top panel**: The calculated mass of the single species, from which the bottom frequency distribution of charged particles arose.

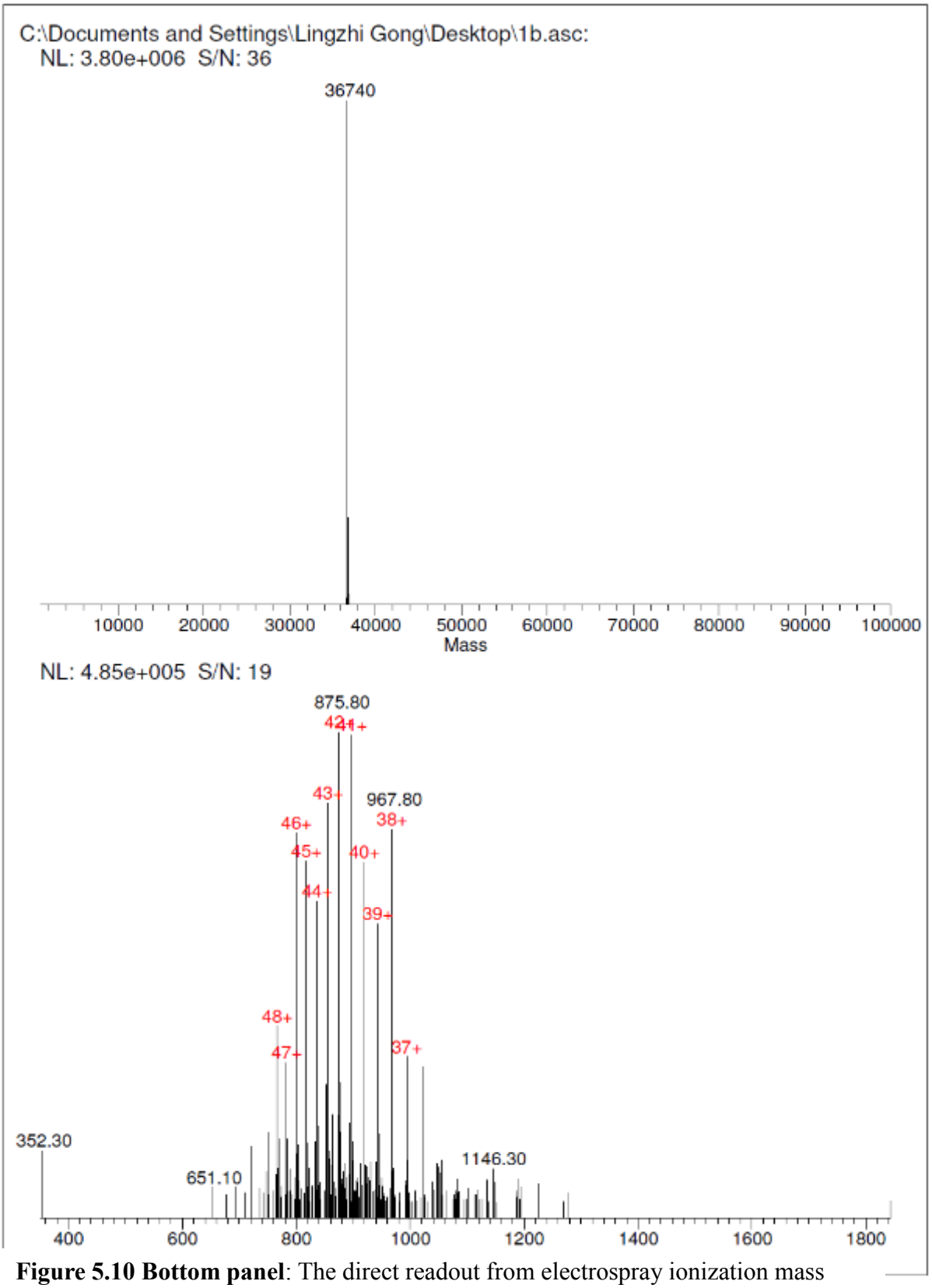


Figure 5.10 Bottom panel: The direct readout from electrospray ionization mass spectrometry of GST-OutB mutant T138C treated with 1,4-bismaleimidobutane (BMB). The *x*-axis is the mass-to-charge- ratio of the ion, and the *y*-axis is the relative intensity of the species. **Top panel**: The calculated mass of the single species, from which the bottom frequency distribution of charged particles arose.

5.7 Conclusions

Cysteine scanning mutagenesis was used to test if the interaction observed between the N0 domain of OutD and the inner membrane protein OutB was similar to the other interactions observed with the N-terminal domain of the secretin.

Interactions previously observed by pull down assays were used to model possible residues, which if the interactions were similar, could kinetically 'trap' the complex by forming an artificial disulphide bond.

Cysteine point mutations were successfully introduced on the 1st beta strand of OutB (residues V136, Y137, T138) and the 1st and 3rd beta strands of the N0 domain (S4, A5, S6 and T63, I64, S65 respectively). The introduction of cysteines on the first beta strand of N0, lead to the formation of homodimers, as assessed by SDS-PAGE analysis. This provided the added complication as N0 homodimers (22 kDA) were similar in size for the anticipated OutB-OutD heterodimers (20 kDa). To overcome this problem, OutB was purified with its N-terminal GST-tag intact, so as to provide a greater difference in mass between N0-N0 dimers and any possible OutB+N0 complexes.

Unfortunately, no increases in mass were visible on SDS-PAGE gels upon addition of an oxidising agent to induce disulphide formation. The addition of crosslinking reagents of different lengths (BMOE: 7Å and BMB:10.9Å) in order to attempt to bridge the gap between any cysteines which were close in space, but not quite close enough to form disulphides, also failed to link the 2 proteins together.

Addition of the crosslinking reagent to the protein under the experimental conditions was checked using GC-MS (gas chromatography, mass spectrometry) in order to check the validity

of the experimental method. Given that an interaction was observed using a pull down assay, it was decided to investigate the interaction further using more sensitive techniques than cysteine scanning mutagenesis.

Chapter 6: Solution NMR studies on the structure of OutB and its interaction with N0

6.1 Aim

To complement X-ray studies of the structure of OutB and investigate its interaction with OutD-N0.

6.2 Introduction

As it did not prove possible to co-purify OutB and OutD-N0, the best available method to determine the atomic details of the interaction between these two proteins was solution NMR.

6.3 Assignment of OutB

A standard triple resonance backbone assignment approach was used to assign the amide protons of OutB 112-220 using HNCACB and CBCACONH spectra. The HNCACB spectra correlates each NH group with the chemical shift of the current residue's CA and CB(strongly) and that of neighboring residue's CA and CBs (weakly). The CBCA(CO)NH spectra only correlates the NH group of the current residue to the preceding CA and CB chemical shifts. By comparing both these spectra, the current residue and its preceding CA / CB resonances can be identified. The succeeding amide proton chemical shift was identified using the chemical shifts of the preceding CA CB to search for the amide giving same chemical shift within the current residue's CA, CB on HNCACB spectra. Proline residues, where amide proton is absent, were not assigned. Assignment of side-chain resonances was also not performed.

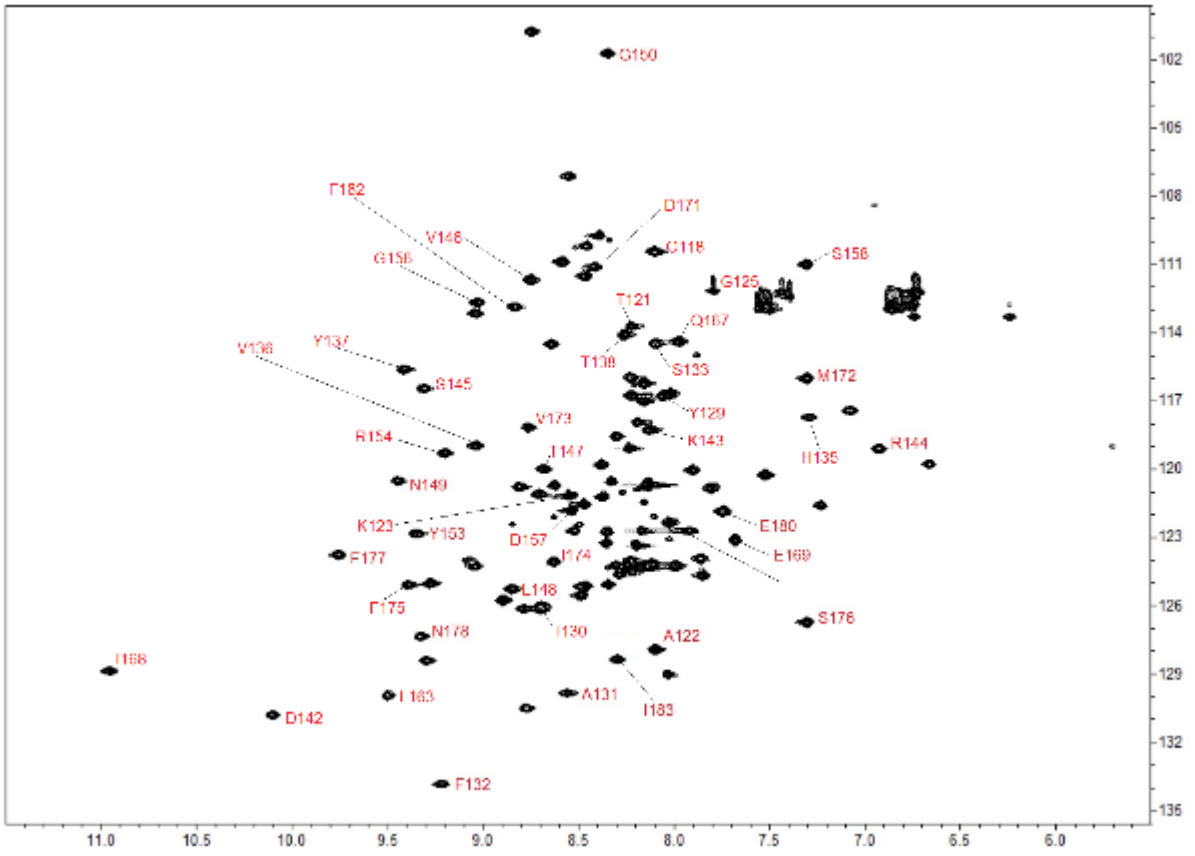


Figure 6.1 ^{15}N HSQC of OutB

The HSQC of OutB was measured and a partial assignment of the N-H amide resonances was achieved using the HNCACB and CBCACONH spectra.

6.4 Interaction of OutB and N0

The interaction initially revealed through pull down assays (Chapter 4) was probed via NMR methods. By labelling each protein in turn and adding unlabeled partner protein, changes in the amide chemical shifts were observed in both experiments (Recording ^{1}H-^{15}N HSQC spectra of an isotopically labeled OutB protein and titrating unlabeled [^{14}N abundant] OutD-N0 and the reverse labelling scheme). In titrations, cross titrations were performed to eliminate the dilution of the labeled protein. In cross titration, two samples were prepared; one sample was labeled protein only and the second sample had the same concentration of labeled protein and a 5 equivalents of unlabeled titration partner. During the titration the concentration of labeled protein did not change, so it was not necessary to correct signal intensities due to dilution.

The interaction between OutB and OutD-N0 appears to be strong. Upon titration of the unlabeled partner protein, a new and well dispersed set of signals were observed in the 15N-HSQC spectrum of both proteins (Figures 6.2 and 6.3). The peaks corresponding to the "free" form became weaker and a new spectrum of peaks corresponding to the "bound" form appear and become more intensified. The co-presence of the "free" and "bound" forms show that the interaction between OutB and OutD-N0 is in "slow exchange" (on the NMR time scale – miliseconds) and thus indicates the tight binding between the two proteins.

Since the titration experiments used 40μM of labeled OutB with addition of 1.3 equivalents of OutD-N0, all the OutB shifted to the "bound" state, indicating that the stoichiometry of the complex would be a 1:1 ratio, with a Kd < 40μM.

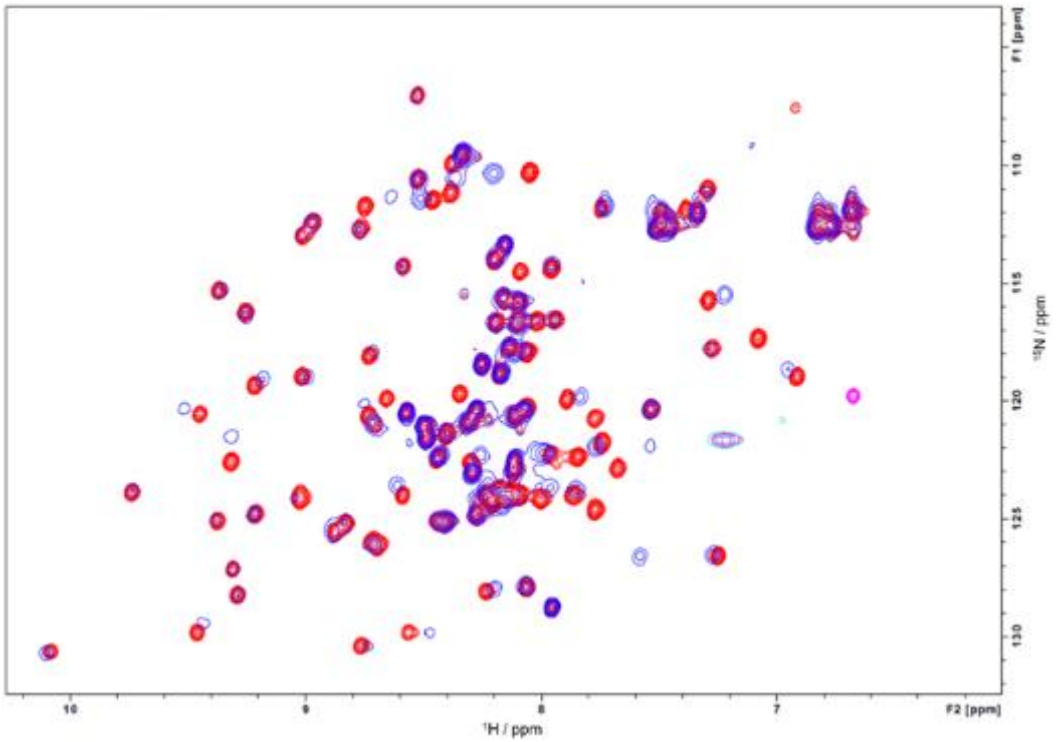


Figure 6.2 ^{15}N HSQC of OutB
The HSQC of OutB was measured in the absence (red) and presence (blue) of 5 molar equivalents of unlabeled OutD-N0.

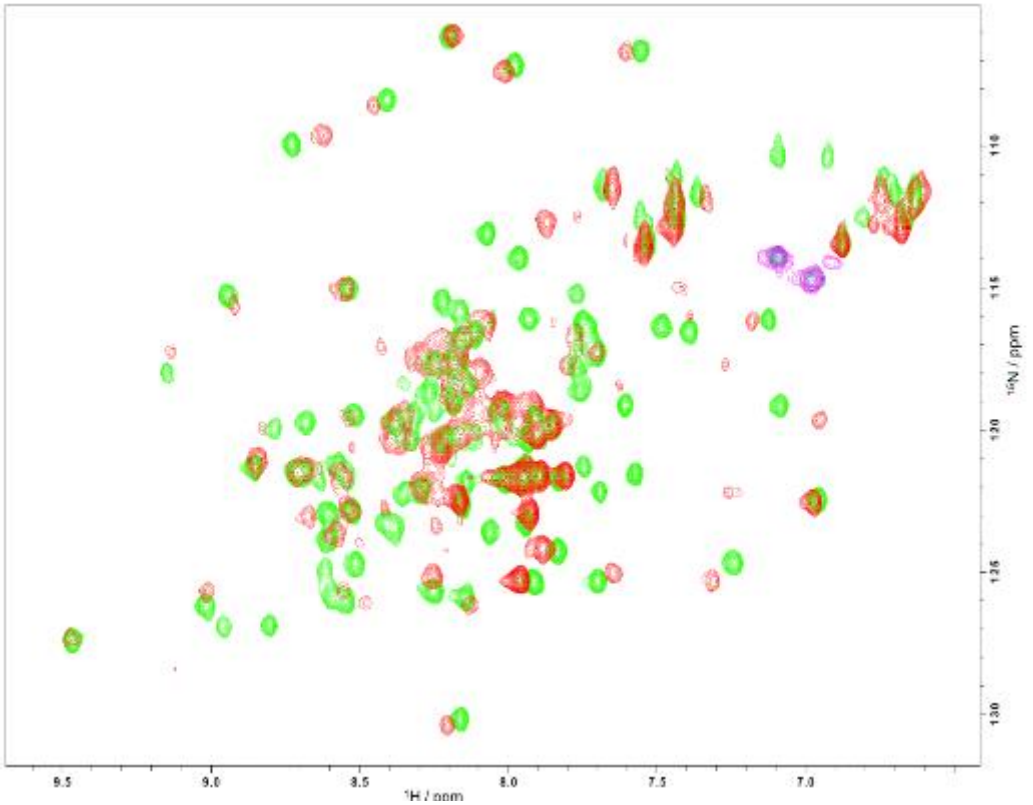


Figure 6.3 ^{15}N HSQC of OutD-N0
The HSQC of OutD-N0 was measured in the absence (green) and presence (red) of 5 molar equivalents of unlabeled OutB.

6.5 Homology modelling of OutD-N0

Sequence identity between OutD-N0 and GspD-N0 is 52%. This indicates that there is a high likelihood of both N0 domains sharing a similar fold. SWISS-MODEL (http://swissmodel.expasy.org/) and MODELLER (https://salilab.org/modeller/) were used to generate homology models of OutD-N0 according to sequence alignment and homology with the GspD-N0 structure (PDB:3EZJ). The structures of OutD-N0 generated by the two different servers are based on the same known structure. The resultant models are therefore very similar with a backbone RMSD of 0.25Å (Figure 6.4). The generated models give some idea of the molecule backbone orientation.

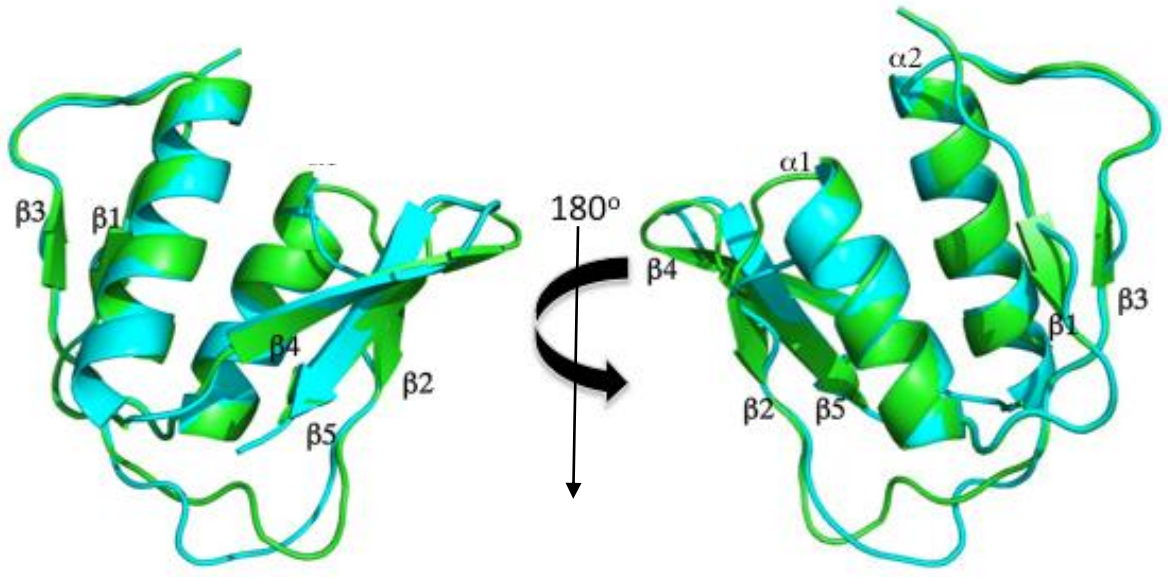


Figure 6.4 Homology based OutD-N0 structure prediction. SWISS-MODEL generated structure in green. MODELLER generated structure is cyan. (Adapted from Gu *et al*., 2011)

6.6 Attempts at modelling the complex based on chemical shift analysis

Both OutB and OutD-N0 showed some peak shifting in their respective HSQC spectra when titrating with the other (Figures 6.2 and 6.3 respectively). This indicates that there are some local structural changes as a result of the two proteins interacting. Some detected peak shifts were greater than others such that the peaks originating from the bound or unbound forms did not overlap.

This indicates a larger degree of local structure perturbation as a result of binding, and it is therefore possible to surmise that these areas are closer to the binding interface, than those residues showing only a small degree of change in their chemical shift values.

As a result, peaks that showed a large degree of chemical shift perturbation, and no longer overlapped once saturated with binding partner, were selected and mapped to the surfaces of their corresponding proteins (Figures 6.5 and 6.6).

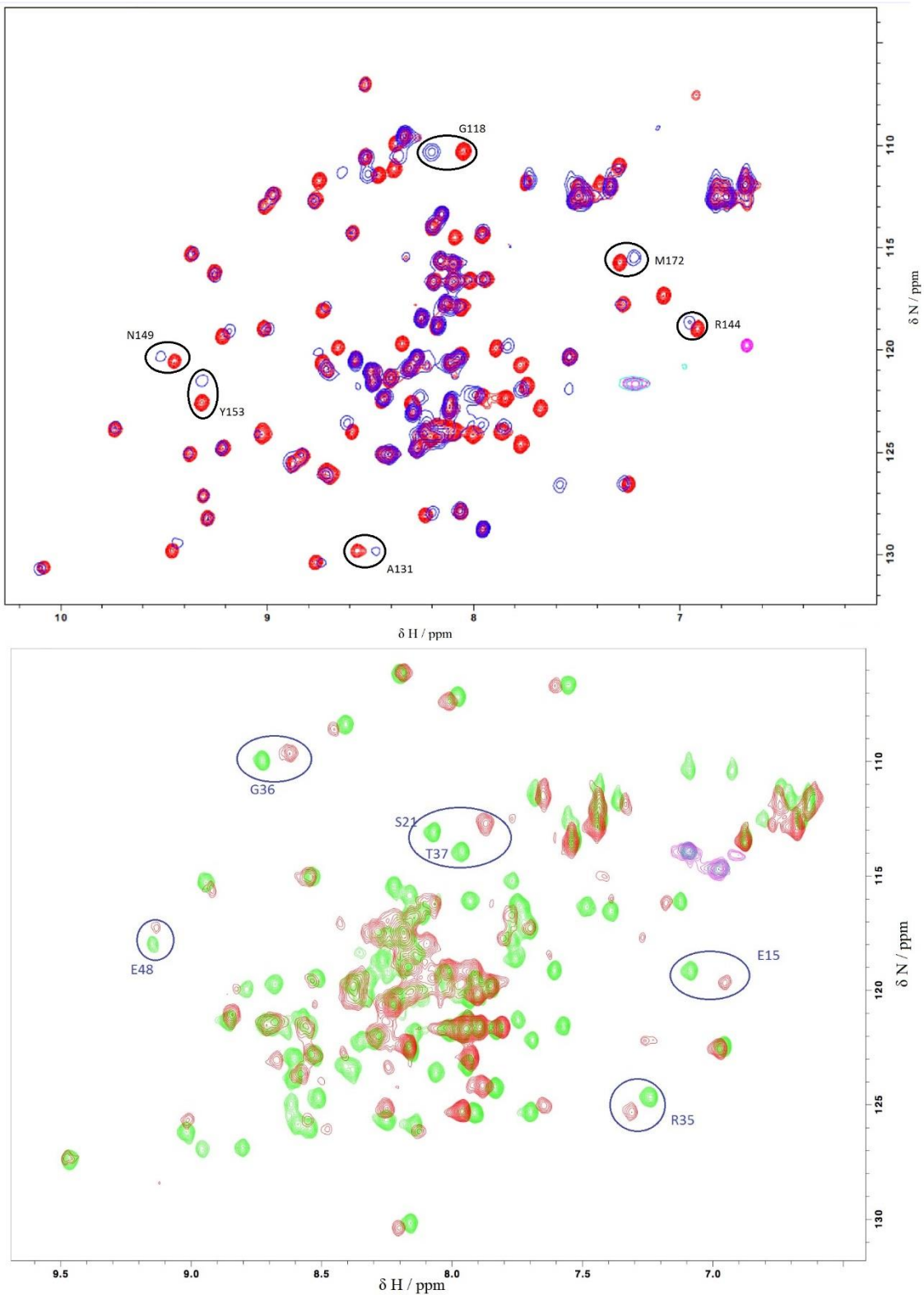


Figure 6.5 Top: ^{15}N-HSQC of OutB in the absence (red) and presence (blue) of unlabelled OutD-N0. Bottom: ^{15}N-HSQC of OutD-N0 in the absence (green) and presence (red) of unlabelled OutD-N0

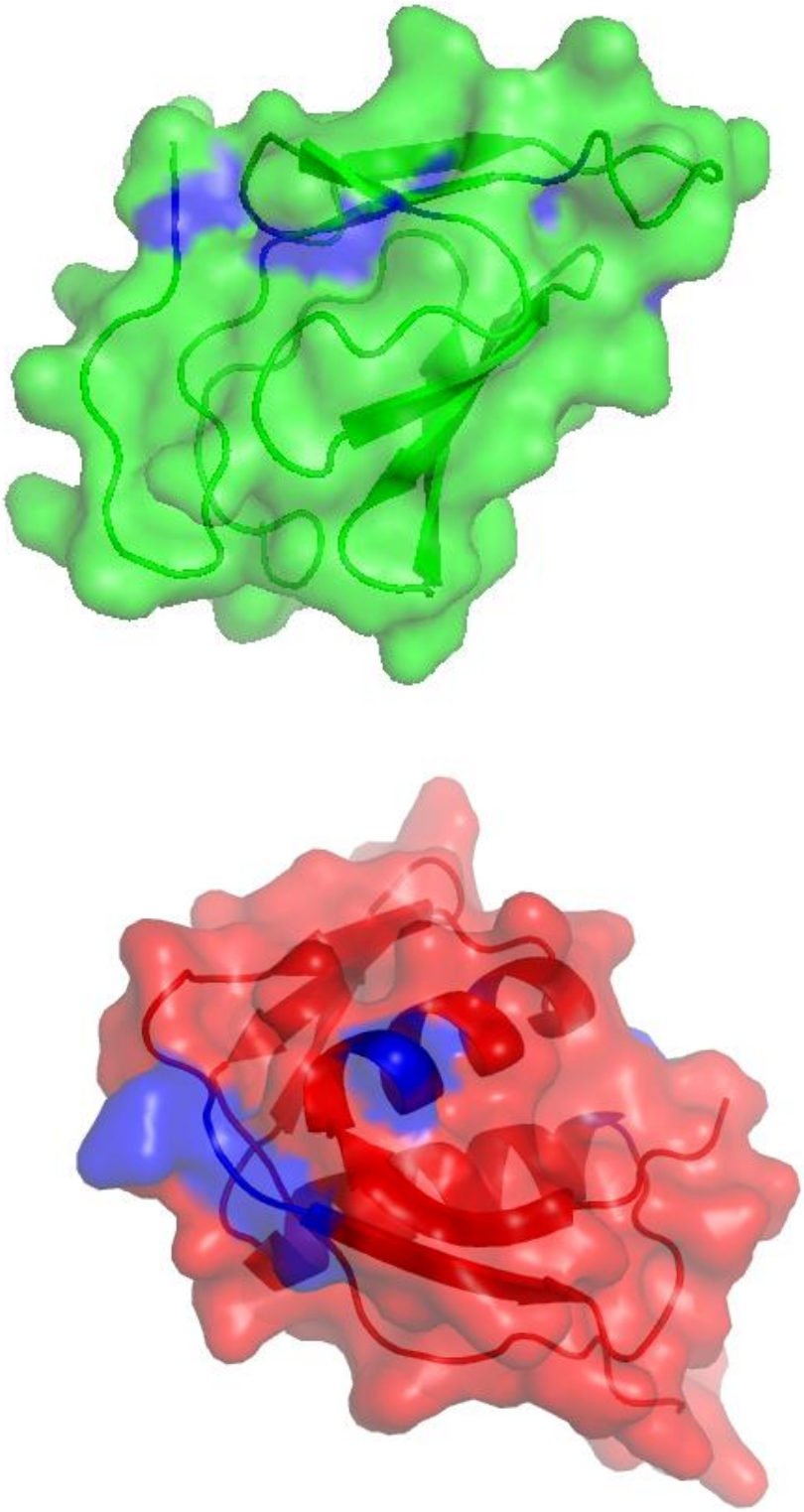


Figure 6.6 Mapping the largest chemical shifts to protein surfaces

Mapping the largest chemical shifts to the surfaces of OutB (top) and OutD-N0 (bottom). The residues circled in Figure 5.5 are shown in blue, with the rest of the protein surface show in either green (OutB) or red (OutD-N0)

Mapping the largest chemical shifts to the surfaces of OutB and OutD-N0 resulted in each case, to a reasonably well defined, discrete area on the surface of each protein: a surface formed by beta-strands 1 through 3 of OutB and a surface proximal to beta-1 of OutD-N0 (Figure 6.6). As a result, it was possible to specify these surfaces to the high ambiguity drive protein-protein docking (HADDOCK) server, in order to model the lowest energy configuration in which the 2 proteins may interact. The resultant model is shown in Figure 6.7 below.

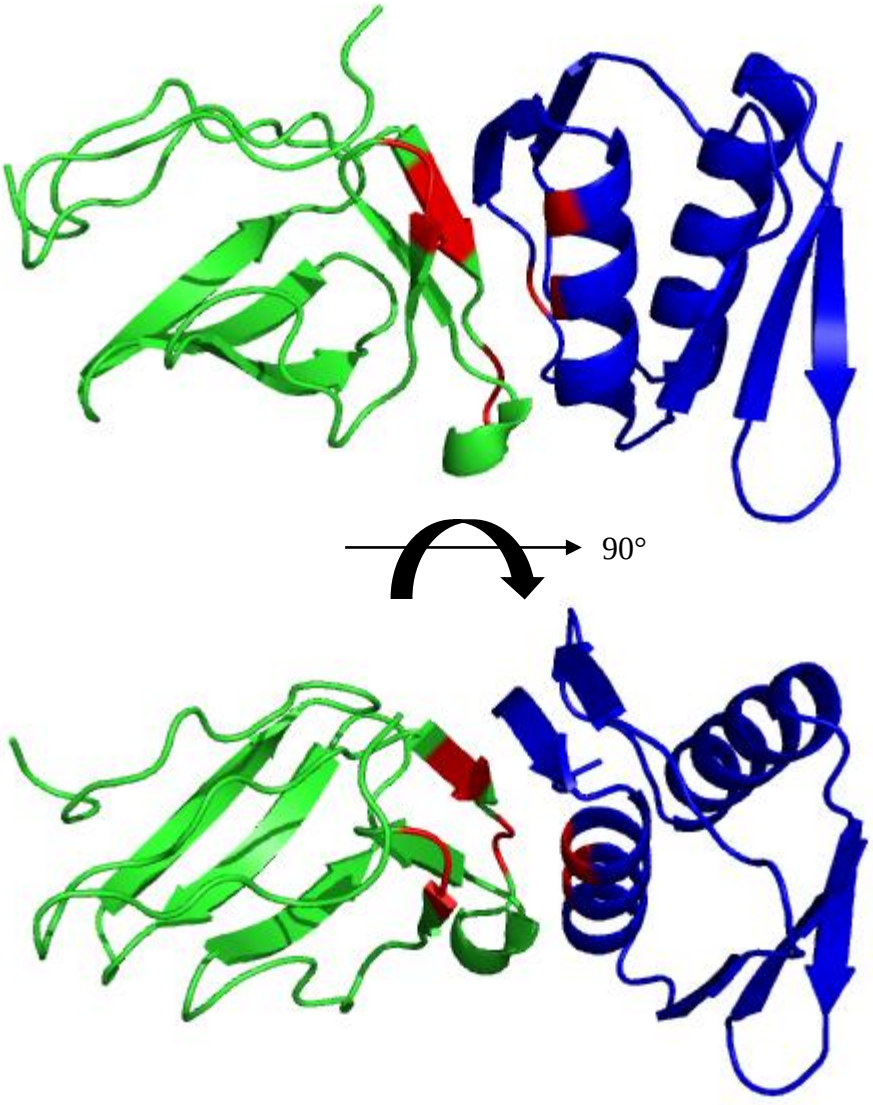


Figure 6.7 Model of the interaction between OutB and OutD-N0 based on the chemical shift analysis outlined above, with OutB in green and OutD-N0 in blue. The residues which showed large chemical shifts and we used as docking parameters are shown in red.

6.7 Filtered NOE Experiments

The Nuclear Overhauser Effect (NOE) can be used to establish if a protein-protein interaction is a specific or non-specific interaction. If 2 atoms, 1 from each partner protein, are continually close together, for a long enough period of time for magnetisation to be transferred between them, this would indicate they are involved in, or near to, a defined binding site. If no such interactions are observed, it is assumed that the changes seen in the HSQC measurements are due to a non-specific or generalised interaction between the two proteins.

To identify only intermolecular NOEs between 2 proteins, and avoid also measuring intramolecular NOEs within the same protein, it is necessary to isotopically label one and not the other. It is then possible for a pulse sequence to omit intramolecular NOEs from within both labeled and unlabeled constituents. The filtered experiments run are designed to detect inter-molecular contacts, using the fact that one component of the complex is labelled and one not.

The pulse sequence starts by eliminating the 1H{15N/13C} magnetisation, while retaining that originating from the unlabeled component. Then 1H chemical shift of this 1H{14N/12C} magnetisation is recorded, and gives one of the indirect dimensions. Then an NOE mixing period allows this magnetisation to transfer over to the labelled component, if the hydrogen in question is close, (<6 angstroms) to the interface. The "acceptor" hydrogen is then measured, along with its attached heteronucleus.

Using 13C / 15N doubly labelled protein-protein, NOEs can be selective observed be so called isotope filtered and edited experiments. In this experiment, OutB was labelled uniformly with 15N and 13C while OutD-N0 is at the natural isotopic abundance.

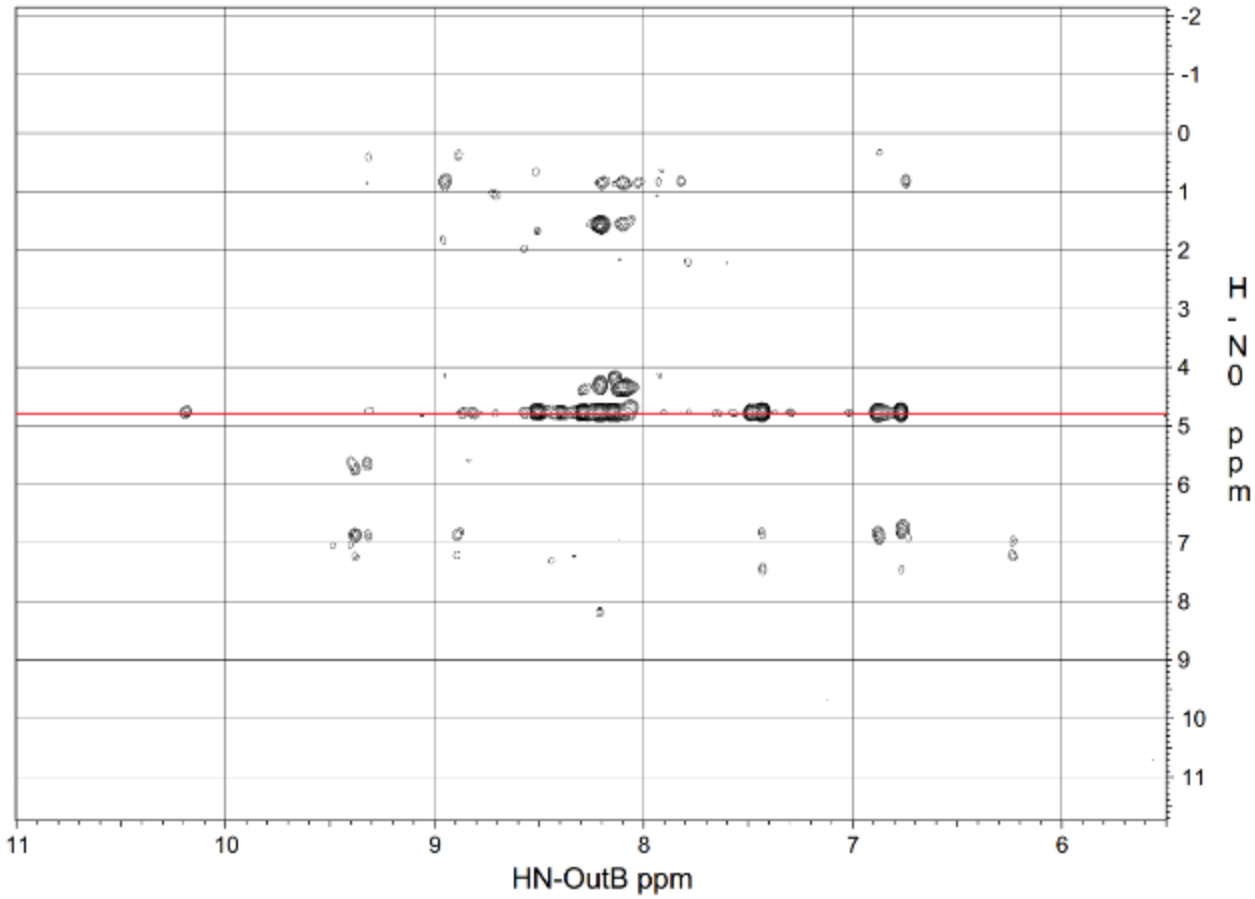


Figure 6.8 Projection of the isotope filtered experiment

OutB was isotopically labelled (15N / 13C) and OutD-N0 was not. The chemical shift of the OutB 'donor' protons is shown along the *x*-axis and the chemical shift of the receiver protons on OutD-N0 on the *y*-axis. Note numerous receiver peaks at 4.7ppm (red line). These are solvent exchange hydrogens where the receiving proton of the observed NOE is water.

The presence of observable NOEs (Figure 6.8) between the 2 proteins in the complex indicates that 2 atoms, 1 from each partner protein, are continually close together, for a long enough period of time for magnetisation to be transferred between them, indicating they are involved in, or near to, a defined binding site.

6.8 Discussion

In this chapter it has been demonstrated that the inner membrane protein OutB of the type 2 secretion system in *Dickeya dadantii* forms a strong interaction with the N-terminal domain (N0) of the outer membrane secretin OutD. This is similar to the GspC proteins in other species.

Enterotoxigenic *E. coli* GspC-HR has a strong interaction with the N0 domain of the secretin, GspD (Korotkov, 2011) in comparison to that observed with the corresponding proteins in *D. dadantii* (Gu et al., 2012a). It is possible that the role of assembling the functional type II secretion system fulfilled by GspC-HR in enterotoxigenic *E. coli* has been partially taken-up by OutB in *D. dadantii* freeing up *D. dadantii* OutC-HR for a greater involvement in other functions, such as determining the specificity of substrates secreted by the system.

Based on the similarity of OutB with various GspC structures, a model was presented on how OutB and OutD may interact. This model was generated by mapping the largest chemical shifts that occur in the HSQC of each respective protein as a result of binding the other, to their respective surfaces and using these residues, which experience the largest chemical shifts, as docking parameters for the modelling programme HADDOCK.

It is entirely plausible, though not demonstrated by experimental means, that the first β-strand of GspB could interact with the first β-strand of GspD-N0; this brings two small hydrophobic

patches together involving the conserved Ala and Val on the surface of β1 of GspB and conserved Phe of GspD-N0.

Calculating the electrostatic surface of the region of GspB that may interact with N0 reveals a central band of positive potential with negative bands either side (Chapter 3, Fig. 3.13, other data not shown). *D. dadantii* and *E. coli* GspC-HR domains have a neutral central band with negative potential either side and are quite similar to one another (Chapter 3, Figure 3.13b, c), while the periplasmic domain of *P. aeruginosa* PilP is more similar to GspB in terms of electrostatic potential, having a pronounced central band of positive potential. The striking similarity of PilP to GspB both structurally and electrostatically suggests they have similar functions in the type II secretion and type IV pili systems respectively.

The results show that OutB and OutD interact. Given that the structures of GspB and *E. coli* GspC-HR are closely similar supports the view that the role of *D. dadantii* GspC-HR in assembling the secretion system has been assumed in part or completely by GspB. This may possibly allow *D. dadantii* GspC-HR to become more specialised in substrate selection. A role which is supported by various studies (Gu *et al* 2012a, Pineau *et al*. 2014).

Chapter 7: Conclusion and future directions

7.1 Work completed

Here, I have made truncations of OutB and successfully shown that its architecture is similar to that of the HR domain of OutC. The truncation OutB 112-202 gave crystals that diffracted well. The structure comprises two three-stranded antiparallel β-sheets; the β-strands sequentially form the up-down-up β-sheets, so the first sheet comprises strands 1, 2 and 3 and the second 4, 5 and 6. The two sheets are at approximately 70° to each other so the structure forms a small barrel with a hydrophobic core. There are three residues in 310-helix conformation in the loop between β1 and β2. Perhaps the most striking feature of the structure is the quantity of irregular polypeptide, between the short β-strands, especially but not limited to the β3/ β4 loop, and at the amino- and carboxy-ends of the small β-barrel. This is a new structure, refined coordinates and observed structure factors are deposited in the PDB with accession code 4WFW.

Pull-down assays had shown that OutB binds to the N-terminal domain (N0) of OutD. Attempts to cross-link OutB and N0 were surprisingly unsuccessful, possibly hampered by the proclivity of N0 to form cross-linked homodimers. Self-association of N0 had not been seen before and the evidence suggests it involves β1- β1 interactions.

NMR analyses suggest that OutD-N0 has a reasonable affinity for OutB (< 40μM) and that the interaction involves association of residues on the first beta-strand of each protein.

Although there are only weakly similar sequence homologs to OutB, analysis of the structure showed that structural homologs of OutB are present in the T2 and T4P secretion systems. Analysis of the surface potentials of these structures using a finite-difference solution of the Poisson-Boltzman equation calculated using APBS software, indicated that although the GspC

proteins are structurally similar to OutB, these 2 proteins although similar to one another, have very different surface charge potentials - unsurprising given the lack of sequence similarity to OutB.

What is surprising however, is that OutB bare remarkable similarity to the inner membrane protein of the *Pseudomonas aeruginosa* T4P system, PilP. PilP is known to interact with the cognate secretin of the T4P system, a similar function to that of OutB. This suggests that the largely similar charge distribution on these two proteins may play a similar role, either in interactions with secretins, or in an unknown role they play.

7.2 Future work directly related to this Thesis

Since its discovery nearly 25 years ago, much work has been performed to try and understand the T2SS. In the past decade, there have been great advances in understanding the molecular structure of individual components of the T2SS. These structures have been solved principally by X-ray crystallography as well as a few domains of component proteins by solution NMR.

In this Thesis I have focussed on elucidating the molecular details of the interaction of the T2SS outer membrane secretin, OutD, with the inner-membrane protein OutB in the bacterium *Dickeya dadantii*. The discovery of the interaction between OutB and OutD, presented in this thesis, has raised numerous questions in terms of the temporal arrangement and order in which these interactions occur. Given that the secretin in *D. dadantii* is now known to interact with no fewer than 3 proteins (Gu *et al.* 2012b, Korotkov *et al.* 2011, this work); the inner membrane platform component OutC, its cognate pilotin OutS and now OutB. Both the latter two of these proteins (OutS and OutB) are implicated in assembly of the T2SS system as a whole.

Given these numerous interactions the next, and technically more challenging task, is to investigate the dynamic interplay between all these interactions. Are there multiple interactions present to allow redundancy in assembling the T2SS, given its importance in virulence and survival of bacteria, or is the function of these multiple components within the T2SS gene cluster more subtle than purely 'assembly'. The biological significance of this interaction in *Dickeya dadantii* has also been discussed.

It is possible that OutB and OutC bind to the secretin at different sites. In order to answer this, it is necessary to interpret the NOE data and determine the residues which are close in space as a result of binding. A full assignment of OutB and OutD-N0 in complex is required. However, it may be quicker to determine the exact nature of the complex if OutB and OutD-N0 domains can be co-crystalised and an X-ray structure determined.

Solving the phase problem of such a structure by molecular replacement should be comparatively straightforward given that structures are known for each of the components; thereby avoiding one of the major bottle necks of X-ray crystallography.

However, there is still limited knowledge of how the components interact and assemble to form a functional system, and even whether the system exists fully assembled, waiting for substrates to secrete, or assembles only when required. However, establishing this interaction still leaves many questions about the functionality of the T2SS unanswered.

The exact role of the GspA / B complex observed in some species and the absence of GspB in other systems still remains an unanswered question. I hypothesis this is possibly due to the diverse thickness of peptidoglycan in different bacterial species, which may necessitate or negate the requirement of GspA to provide energy to the process performed by GspB.The absence of GspB homologues I hypothesis is to do with the propensity of the secretin to self

multimerize, so that in those species where the cognate secretin has a high self-affinity, a GspB homologue is not required to organise the secretin correctly.

7.3 Other unanswered questions related to the T2SS

7.3.1 Cryogenic Electron Microscopy & single particle and tomography work could be used to assess the way the T2SS nanomachine is organized in detail

A major advantage of cryoelectron microscopy is that specimens can be observed in a more native, physiological condition - without having to be stained or fixed in any way. This is in stark contrast to X-ray crystallography, which requires crystallisation of the protein(s) under highly artificial conditions, which can lead to functionally irrelevant conformational changes.

Transmission electron microscopy has been successfully used to image entire, isolated T3SS needle complexes from *S. typhimurium* (Schraidt *et al.* 2010).

Cryoelectron microscopy, could if applied successfully to the T2SS, yield important results elucidating the exact arrangement of the inner membrane platform, its stoichiometric arrangement and the exact nature of the associated between the ATPase, IM platform and pseudopilins. This would be essential knowledge, in order to shed light on how energy transfer occurs throughout the system.

7.3.2 The Export Mechanism and energy transduction is poorly understood

At the moment the only known ATPase, associated with the T2SS, are GspE homologs. These proteins are associated with the cytoplasmic side of the inner membrane and it is unclear how

energy derived from ATP hydrolysis on one side of the membrane can be transferred (presumably via the inner membrane platform proteins) to drive pseudopilus assembly in the periplasmic space, and secretion through the outer membrane secretin pore.

Although there have been great advances in understanding the molecular structure of individual components of the T2SS, most of these new structures have been of soluble domains from constituent proteins not the full length proteins. This is understandable given the current difficulties in handling and working with membrane proteins. However, given that an intrinsic purpose of this system is to span 2 membranes, it will be necessary to establish the structure of full length proteins in order to understand how the numerous proteins of the inner membrane interact and pack.

The IMP is much more heterogeneous than both the outer membrane pore (homo-multimeric), and pseudopilus, (consisting of major and minor pseudopilins) consisting of at least 5 proteins (homologs of GspF, L, M, C and E). I have begun work on cloning and purifying the GspF homolog from *D. dadantii*, OutF (Appendix G). This protein is the only protein of the inner membrane platform (IMP) to possess more than 1 transmembrane helix, and is therefore presumed to form the 'core' or centre of an IMP complex, given its presumably inherent greater stability within the membrane. Determining a structure of the integral membrane proteins of the IMP will be essential in further understanding the structure and function of the T2SS and its role in pathogenesis.

7.3.3 Exoprotein Selection

A fundamental problem related to the T2SS, and many other bacterial secretion systems, is how fully folded proteins are distinguished from those that are not secreted. Proteins routed through the T2SS are known to acquire a highly ordered structure in the periplasmic space before translocation across the outer membrane and release to the environment (Hirst and Holmgren, 1987; Pugsley, 1992; Poquet et al., 1993). Consequently, they are expected to contain a targeting motif in their folded structure, as no linear, amino acid signal has been detected.

In several studies, the information needed for active extracellular targeting has been located to segments covering 60–80 amino acids of the mature protein (Lu and Lory, 1996; Sauvonnet and Pugsley, 1996; Lindeberg et al., 1998, Palomaki *et al.* 2002). But in no cases have the underlying recognition motifs been unveiled. It is hypothesised that the actual targeting motifs in even these proteins should require a far smaller number, perhaps <10, residues than the 60–80 amino acids present in the primary sequence of the secretion promoting segments (Palomaki *et al.* 2002).

A study using chimeras made between the PelC protein of *E. chrysanthemi* and the PelL of *E. carotovora* (Lindeberg et al., 1998), showed a single C-terminal region composed of loops as a primary source of species-specific targeting information of PelC. Two helical regions also reported to be required for secretion were suggested to be needed for the proper positioning of the loop region, but results from this an other studies are still inconclusive as to the potential nature of a 3D structural motif that may target proteins for secretion by the T2SS.

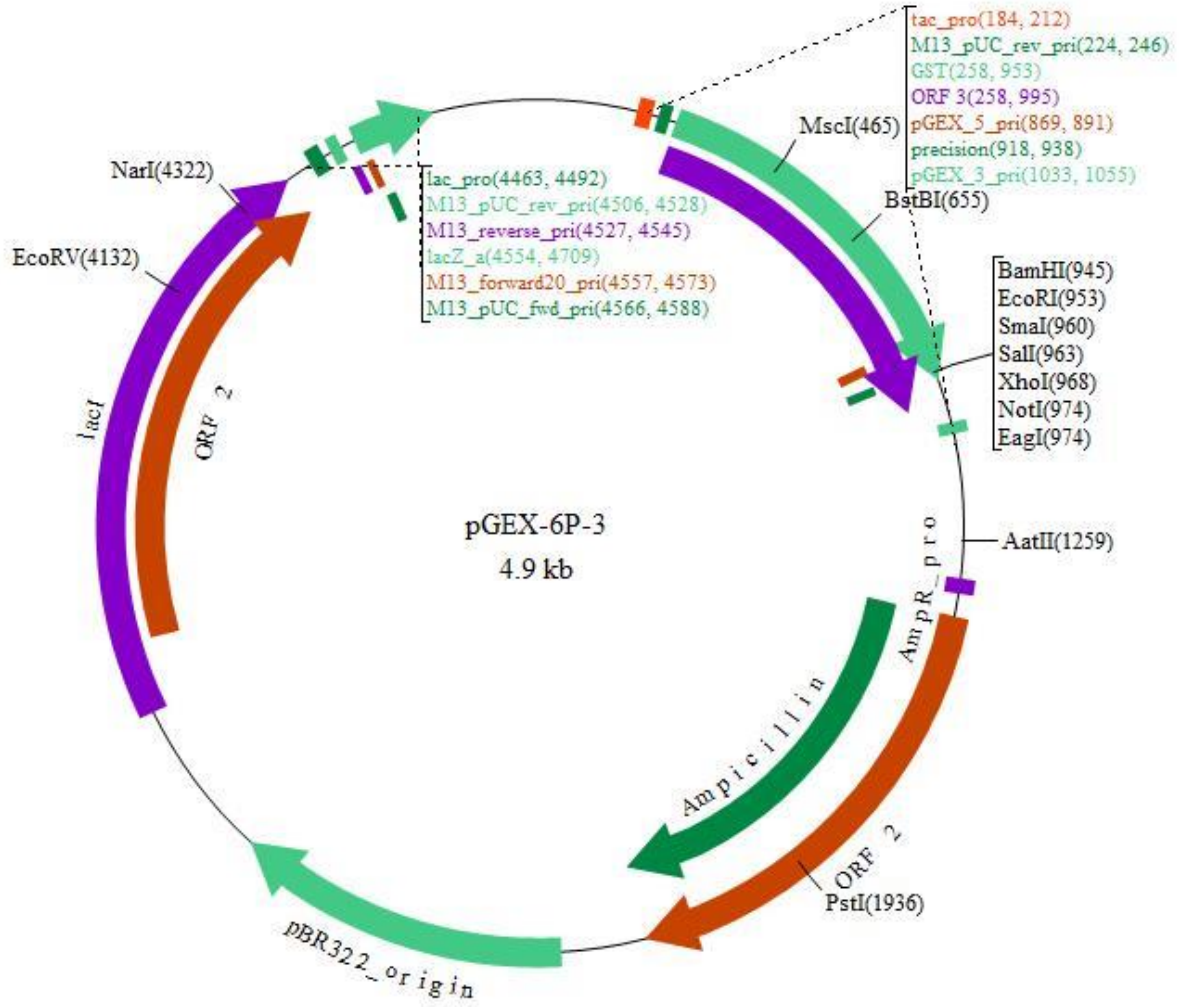
7.3.4 Concluding Remarks

As more information is collected, the organization of the T2S and T4P machineries will become ever clearer, and it will be possible to understand in detail the precise mechanisms by which these two related systems function. TO understand fully the processes through which these systems are regulated, the way in which mechanical energy is transduced from cytoplasmic ATPase(s) across the inner membrane to assemble components in the periplasm, & the selection of exoprotein substrates are all areas in this field of research that require further research.

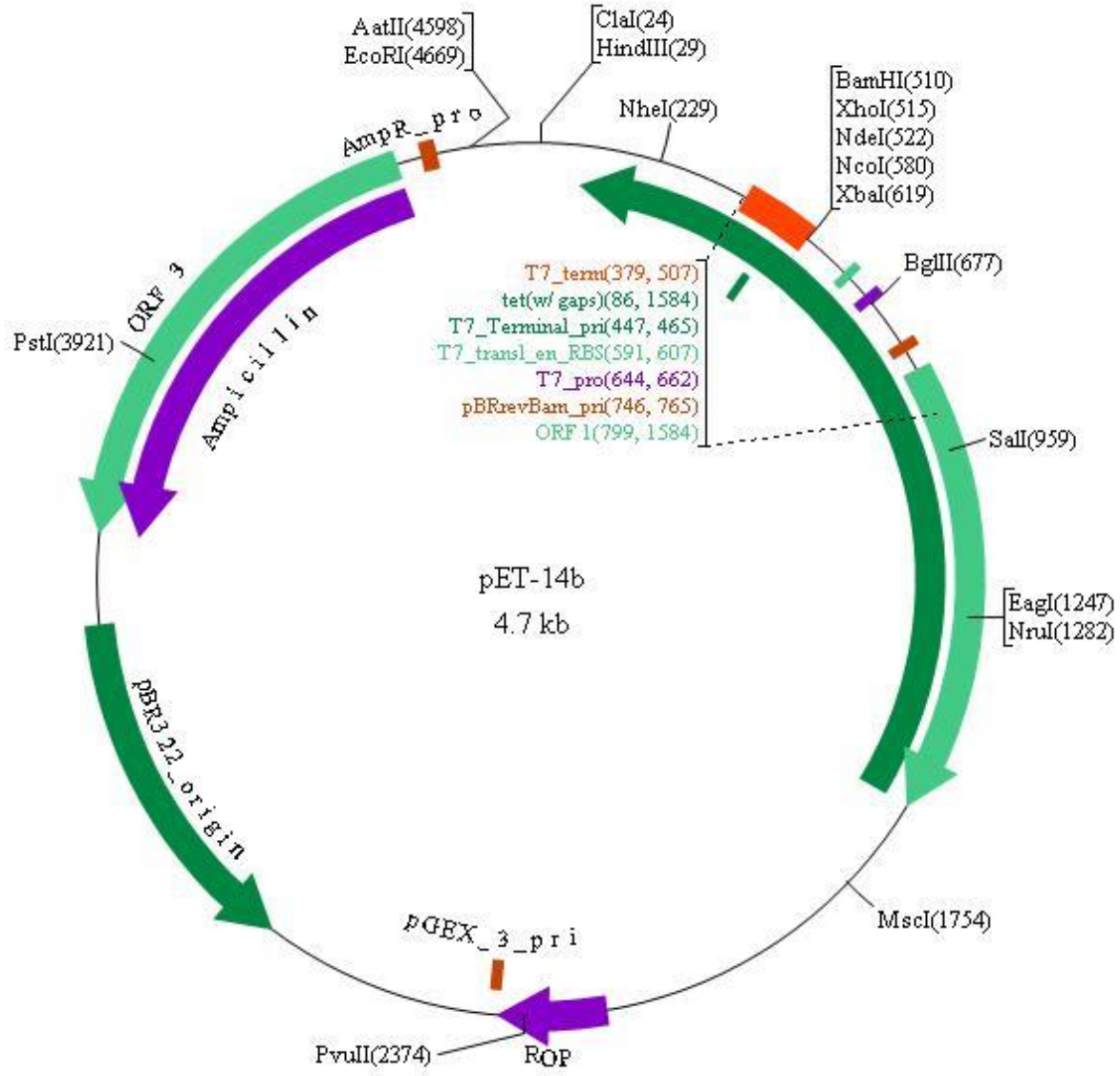
Studies in the T2SS have already started to shed light on exoprotein selection, showing that the N-terminal domains (NTD) of the secretin in *K. oxytoca* could be exchanged with that of *Erwinia chrysanthemi* with no defect in *Klebsiella*-specific protein secretion (Guilvout *et al.* 1999). By contrast, a direct interaction between the secretin with secreted substrates was observed in *E. chrysanthemi*, suggesting that this N-terminal domain plays a role in *Erwinia*-specific secretion (Bouley *et al.* 2001). Answering these questions will help to generate a more complete model of how both the T2S andT4P systems function.

Appendix

(A) 1: pGEX-6P-3 Plasmid



(A) 2: pET-14b Plasmid



(B) Mutant primers

eurofins mwg|operon

Oligonucleotide Synthesis Report

Page 1/1

Mr. Piers Rycroft
Queen Mary University of London
SBCS

Order ID: 2886943
Customer ID: 20876
Your Order ID (PO#): 9233367

Order Date: 14/01/2013
Lab No.: 1241
No. of Oligos: 2/2

Eurofins MWG Operon
Anzingerstraße 7a
D- 85560 Ebersberg

No.	Oligo Name	Sequence (5' -> 3')	Yield [OD]	Yield [µg]	Yield [nmol]	Concentration [pmol/µl]	Vol. for 100pmol/µl	Tm [°C]	MW [g/mol]	GC-Content	Synthesis Scale	Purification	Modification	Barcode IDO	QC Report
1	OutB 193 Fwd	cgttgcaggattggccgggcTAAa aaccgggcgatgacgccg (42)	8.4	236	18.1	-	181	> 75	13030	64.3 %	0.01 µmol	HPSF	-	015672986	-
2	OutB 193 Rev	CggcgtcatcgcccggtttTTAgcc cggccaatcctgcaacg (42)	6.1	184	14.4	-	144	> 75	12803	64.3 %	0.01 µmol	HPSF	-	015672987	-

Eurofins MWG Operon is certified according to ISO 9001:2000

Customer Support: support-eu@eurofins.com | +49 8092 8289-77

eurofins mwg|operon

Oligonucleotide Synthesis Report

Page 1/1

Mr. Piers Rycroft
Queen Mary University of London
SBCS

Order ID: 2882813
Customer ID: 20876
Your Order ID (PO#): 9232531

Order Date: 08/01/2013
Lab No.: 2583
No. of Oligos: 2/2

Eurofins MWG Operon
Anzingerstraße 7a
D- 85560 Ebersberg

No.	Oligo Name	Sequence (5' -> 3')	Yield [OD]	Yield [µg]	Yield [nmol]	Concentration [pmol/µl]	Vol. for 100pmol/µl	Tm [°C]	MW [g/mol]	GC-Content	Synthesis Scale	Purification	Modification	Barcode IDO	QC Report
1	OutB 202 Fwd	CCG GGC GAT GAC GCC GCG CAA TAA AAT GAG CAG GAG CCG ACG (42)	11.0	300	23.0	-	230	> 75	13017	64.3 %	0.01 µmol	HPSF	-	015646851	-
2	OutB 202 Rev	CGT CGG CTC CTG CTC ATT TTA TTG CGC GGC GTC ATC GCC CGG (42)	8.1	254	19.8	-	198	> 75	12816	64.3 %	0.01 µmol	HPSF	-	015646852	-

Eurofins MWG Operon is certified according to ISO 9001:2000

Customer Support: support-eu@eurofins.com | +49 8092 8289-77

(C) X-ray data sets

Summary data for Project: mx7197v39 Crystal: x2681collectM4S11 Dataset: SAD

	Overall	**InnerShell**	**OuterShell**
Low resolution limit	36.33	36.33	2.21
High resolution limit	2.15	9.62	2.15
Rmerge (within I+/I-)	0.038	0.023	0.574
Rmerge (all I+ and I-)	0.040	0.024	0.585
Rmeas (within I+/I-)	0.043	0.027	0.710
Rmeas (all I+ & I-)	0.043	0.027	0.660
Rpim (within I+/I-)	0.020	0.012	0.411
Rpim (all I+ & I-)	0.016	0.010	0.298
Rmerge in top intensity bin	0.024	**-**	**-**
Total number of observations	37835	453	1555
Total number unique	5352	83	360
Mean((I)/sd(I))	25.9	60.3	2.5
Mn(I) half-set correlation CC(1/2)	0.999	0.999	0.818
Completeness	99.6	99.3	97.6
Multiplicity	7.1	5.5	4.3
Anomalous completeness	98.4	100.0	89.8
Anomalous multiplicity	4.0	4.6	2.2
DelAnom correlation between half-sets	-0.004	-0.039	0.007
Mid-Slope of Anom Normal Probability	0.936	**-**	

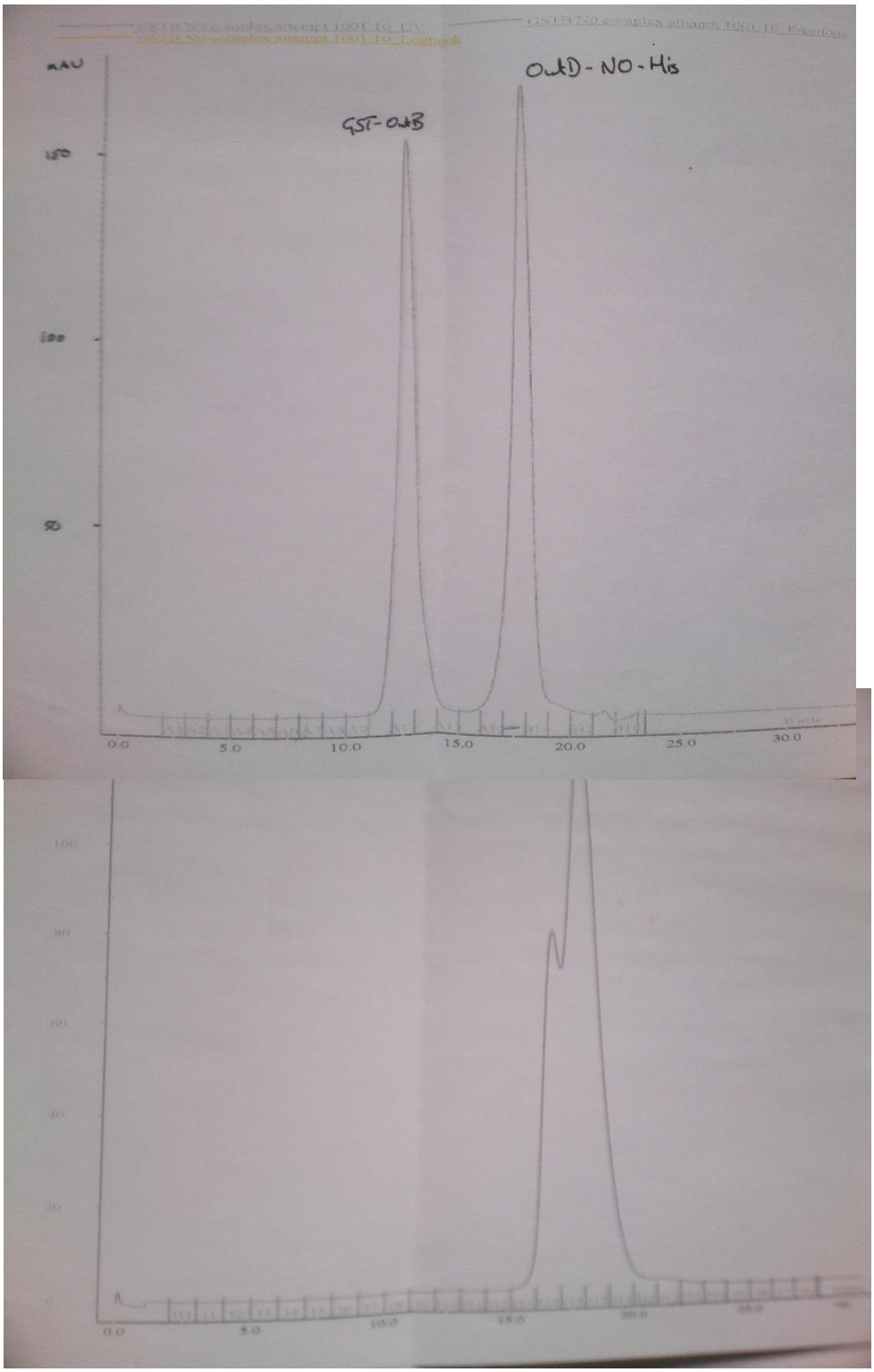
Summary data for Project: mx7197v39 Crystal: x2682collectM4S21 Dataset: SAD

	Overall	**InnerShell**	**OuterShell**
Low resolution limit	36.31	36.31	2.15
High resolution limit	2.10	9.39	2.10
Rmerge (within I+/I-)	0.039	0.029	0.640
Rmerge (all I+ and I-)	0.042	0.030	0.683
Rmeas (within I+/I-)	0.046	0.035	0.761
Rmeas (all I+ & I-)	0.046	0.034	0.747
Rpim (within I+/I-)	0.025	0.019	0.406
Rpim (all I+ & I-)	0.019	0.016	0.298
Rmerge in top intensity bin	0.026	**-**	**-**
Total number of observations	32594	327	2494
Total number unique	5622	92	407
Mean((I)/sd(I))	19.8	38.6	2.6
Mn(I) half-set correlation CC(1/2)	1.000	1.000	0.846
Completeness	99.8	98.8	99.9
Multiplicity	5.8	3.6	6.1
Anomalous completeness	98.3	100.0	99.1
Anomalous multiplicity	3.2	3.6	3.2
DelAnom correlation between half-sets	-0.029	-0.254	-0.040
Mid-Slope of Anom Normal Probability	1.023	**-**	**-**

Summary data for Project: mx7197v39 Crystal: x26812collectM4S121 Dataset: SAD

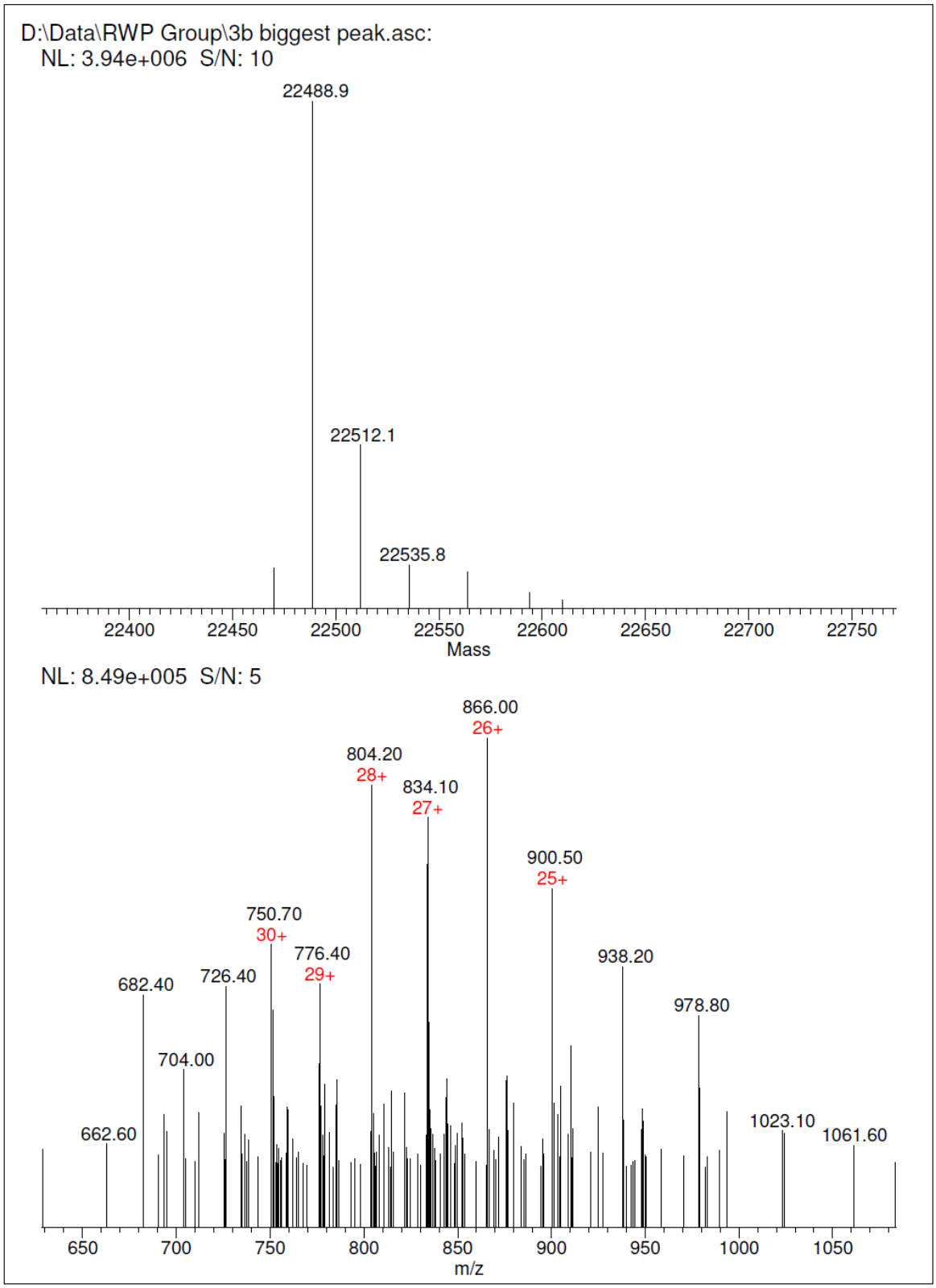
	Overall	**InnerShell**	**OuterShell**
Low resolution limit	36.31	36.31	1.99
High resolution limit	1.94	8.68	1.94
Rmerge (within I+/I-)	0.029	0.027	0.487
Rmerge (all I+ and I-)	0.030	0.029	0.525
Rmeas (within I+/I-)	0.032	0.030	0.583
Rmeas (all I+ & I-)	0.032	0.031	0.584
Rpim (within I+/I-)	0.012	0.012	0.315
Rpim (all I+ & I-)	0.010	0.011	0.248
Rmerge in top intensity bin	0.022	**-**	**-**
Total number of observations	72326	824	2402
Total number unique	6947	114	495
Mean((I)/sd(I))	35.3	70.9	2.6
Mn(I) half-set correlation CC(1/2)	0.999	0.999	0.915
Completeness	99.4	99.8	99.0
Multiplicity	10.4	7.2	4.9
Anomalous completeness	98.7	100.0	91.7
Anomalous multiplicity	5.8	6.5	2.4
DelAnom correlation between half-sets	0.474	0.659	0.081
Mid-Slope of Anom Normal Probability	1.020	**-**	**-**

(D) Size exclusion of B + N0



(E) OutD-N0 Mass spectrometry results

MagTran 1.0 - 04/10/15



(F) OutF Cloning Work – Shevchik laboratory

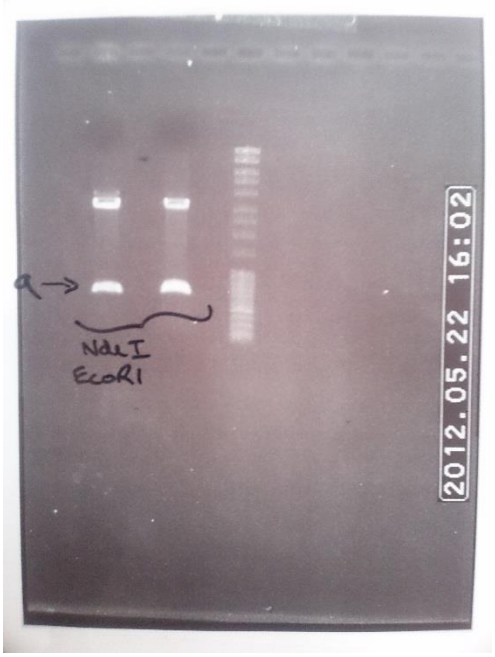


Figure 1 Digestion of pGEX-6P-3 with Nde1 and EcoR1

The OutF gene was to be cloned into the pGEX-6P-3 plasmid for expression. However, the intended vector already contained an insert at the expression site. So the existing insert was removed by double digestion with EcoR1 and Nde1 restriction endonucleases, shown in Figure 2 above; *a* indicates the old insert, no longer required.

Figure 2 PCR amplification and ligation

The OutF gene was amplified by PCR using appropriately designed primers (Lanes 1-4 right hand side)**.**

Ligation of the PCR product into pGEX-6P-3 vector is show left hand side lanes 1-6), a band of excess PCR product is still faintly visible.

Double digested pGEX-6P-3 plasmid s shown in middle to the left of the marker

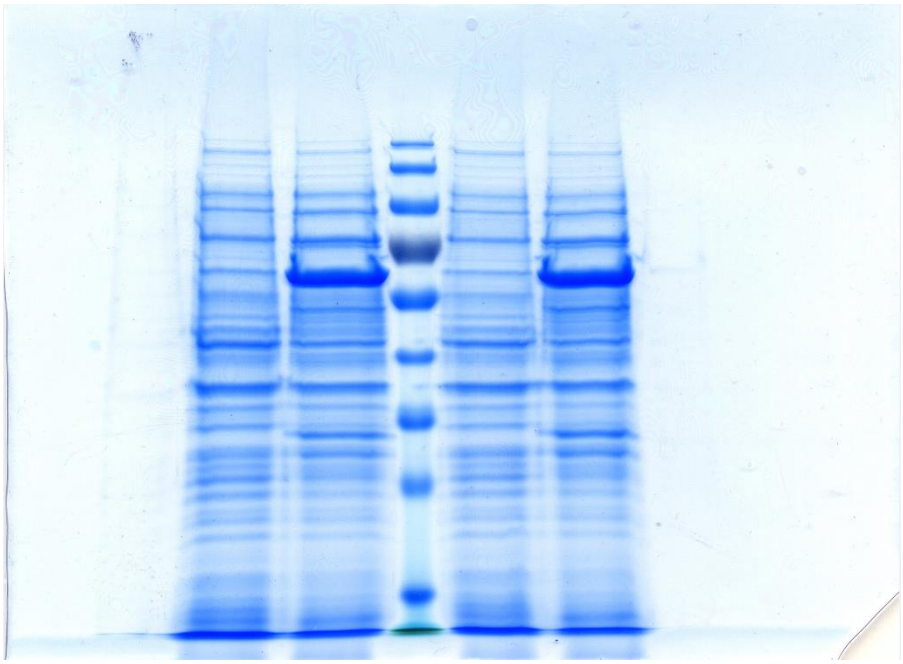


Figure 3 Expression of pGEX-6P-3:OutF

The ligated constructed was the transformed into *E. coli* strain CD43(DE3) for expression. CD43(DE3) is a derivative of BL21(DE3) but is reported to overproduce transmembrane proteins with less toxicity (Miroux et al 1996).

References

Abendroth, J. et al., 2009. The three-dimensional structure of the cytoplasmic domains of EpsF from the type 2 secretion system of Vibrio cholerae. *Journal of Structural Biology*, 166(3), pp.303–315.

Abendroth, J. et al., 2005. The X-ray structure of the type II secretion system complex formed by the N-terminal domain of EpsE and the cytoplasmic domain of EpsL of Vibrio cholerae. *Journal of molecular biology*, 348(4), pp.845–855.

Adams, P.D. et al., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography*, 66 (2), pp.213–221.

Ast, V.M. et al., 2002. Expression of the ExeAB complex of Aeromonas hydrophila is required for the localization and assembly of the ExeD secretion port multimer. *Molecular microbiology*, 44 (1), pp.217–231.

Ayers M., Howell P., Burrows L., 2010. Architecture of the type II secretion and type IV pilus machineries. *Future Microbiol.* 5 (8), pp.1203-1218.

Bally, M. et al., 1992. Protein secretion in Pseudomonas aeruginosa: characterization of seven xcp genes and processing of secretory apparatus components by prepilin peptidase. *Molecular microbiology*, 6 (9), pp.1121–1131.

Bouley, J., Condemine, G. & Shevchik, V.E., 2001. The PDZ domain of OutC and the N-terminal region of OutD determine the secretion specificity of the type II out pathway of Erwinia chrysanthemi. *Journal of molecular biology*, 308 (2), pp.205–219.

Brahms S, Brahms J., 1980. Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *J Mol Biol.* 138 (2), pp.149-178.

Camberg, J.L. et al., 2007. Synergistic stimulation of EpsE ATP hydrolysis by EpsL and acidic phospholipids. *The EMBO journal*, 26(1), pp.19–27.

Camberg, J.L. & Sandkvist, M., 2005. Molecular analysis of the Vibrio cholerae type II secretion ATPase EpsE. *Journal of bacteriology*, 187(1), pp.249–256.

Campos, M. et al., 2013. The type II secretion system - a dynamic fiber assembly nanomachine. *Research in microbiology*, 164(6), pp.545–555.

Chami, M. et al., 2005. Structural insights into the secretin PulD and its trypsin-resistant core. *The Journal of biological chemistry*, 280(45), pp.37732–37741.

Cianciotto, N.P., 2005. Type II secretion: a protein secretion system for all seasons. *Trends in microbiology*, 13 (12), pp.581–588.

Collin S., Guilvout I., Nickerson N., Pugsley A., 2011. Sorting of an integral outer membrane protein via the lipoprotein-specific Lol pathway and a dedicated lipoprotein pilotin. Mol Microbiol. 80 (3), pp.655-665.

Condemine, G. & Shevchik, V.E., 2000. Overproduction of the secretin OutD suppresses the secretion defect of an Erwinia chrysanthemi outB mutant. *Microbiology (Reading, England)*, 146 (3), pp.639–647.

D'Enfert, C. et al., 1989. Protein secretion by gram-negative bacteria. Characterization of two membrane proteins required for pullulanase secretion by Escherichia coli K-12. *The Journal of biological chemistry*, 264 (29), pp.17462–17468.

D'Enfert, C. & Pugsley, A.P., 1989. Klebsiella pneumoniae pulS gene encodes an outer membrane lipoprotein required for pullulanase secretion. *Journal of bacteriology*, 171 (7), pp.3673–3679.

Dalbey R., Wang P., Kuhn A., 2011. Assembly of bacterial inner membrane proteins. *Annu Rev Biochem.* 80, pp.161–187.

Dessvaux M., Hébraud M., Talon R., Henderson I., 2009. Secretion and subcellular localizations of bacterial proteins: a semantic awareness issue. *Trends Microbiol.* 17 (4), pp.139-145.

Dolinsky, T.J. et al., 2007. PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic acids research*, 35, pp.522–525.

Douzi, B. et al., 2011. Deciphering the Xcp Pseudomonas aeruginosa type II secretion machinery through multiple interactions with substrates. *The Journal of biological chemistry*, 286 (47), pp.40792–40801.

Dunstan, R.A. et al., 2013. Assembly of the type II secretion system such as found in Vibrio cholerae depends on the novel Pilotin AspS. *PLoS pathogens*, 9 (1), p.e1003117.

Durand, E. et al., 2005. XcpX controls biogenesis of the Pseudomonas aeruginosa XcpT-containing pseudopilus. *The Journal of biological chemistry*, 280 (36), pp.31378–31389.

Emsley, P. et al., 2010. Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography*, 66 (4), pp.486–501.

Evans, P., 2006. Scaling and assessment of data quality. *Acta crystallographica. Section D, Biological crystallography*, 62 (1), pp.72–82.

Facey S., Kuhn A., 2010. Biogenesis of bacterial inner-membrane proteins. *Cell Mol Life Sci.* 67 (14), pp.2343-2362.

Farizo, K. M., Fiddner, S., Cheung, A. M. & Burns, D. L., 2002. Membrane localization of the S1 subunit of pertussis toxin in Bordetella pertussis and implications for pertussis toxin secretion. *Infect. Immun.* 70, pp.1193–1201.

Fröbel J., Rose P., Müller M., 2011. Early contacts between substrate proteins and TatA translocase component in twin-arginine translocation. *J Biol Chem.* 23, 286 (51), pp.43679-43689.

Goujon, M. et al., 2010. A new bioinformatics analysis tools framework at EMBL-EBI. *Nucleic acids research*, 38, pp.695–699.

Gray, M.D. et al., 2011. In vivo cross-linking of EpsG to EpsL suggests a role for EpsL as an ATPase-pseudopilin coupling protein in the Type II secretion system of Vibrio cholerae. *Molecular microbiology*, 79 (3), pp.786–798.

Grys, T.E. et al., 2005. The StcE protease contributes to intimate adherence of enterohemorrhagic Escherichia coli O157:H7 to host cells. *Infection and immunity*, 73 (3), pp.1295–1303.

Gu, S., Kelly, G., et al., 2012a. Solution structure of homology region (HR) domain of type II secretion system. *The Journal of biological chemistry*, 287 (12), pp.9072–9080.

Gu, S., Rehman, S., et al., 2012b. Structural and functional insights into the pilotin-secretin complex of the type II secretion system. *PLoS pathogens*, 8 (2), p.e1002531.

Guilvout I, Hardie KR, Sauvonnet N, Pugsley A, 1999. Genetic dissection of the outer membrane secretin PulD: are there distinct domains for multimerization and secretion specificity? *J. Bacteriol.* 181, pp.7212–7220.

Guilvout I., Nickerson N., Chami M., Pugsley A.P., 2011. Multimerization-defective ts of dodecameric secretin PulD. *Res Microbiol.* 162 (2), pp.180-190.

Gupta R. S., 2011. Origin of diderm (Gram-negative) bacteria: antibiotic selection pressure rather than endosymbiosis likely led to the evolution of bacterial cells with two membranes *Antonie van Leeuwenhoek*, 100, pp.171–182.

Gupta R. S., 1998. Protein phylogenies and signature sequences: A reappraisal of evolutionary relationships among archaebacteria, eubacteria, and eukaryotes. *Microbiol Mol Biol Rev*, 62 (4), pp.1435–1491.

Hachani, A., Lossi, N. S., Hamilton, A., Jones, C., Bleves, S., Albesa-Jové, D., & Filloux, A. 2011. Type VI Secretion System in Pseudomonas aeruginosa: Secretion and multimerisation of VgrG proteins. *The Journal of Biological Chemistry*, 286 (14), pp.12317–12327.

Hardie K.R., Seydel A., Guilvout I., Pugsley A.P., 1996. The secretin-specific, chaperone-like protein of the general secretory pathway: separation of proteolytic protection and piloting functions. *Mol Microbiol.* 22 (5), pp.967-976.

Hauser, A.R., 2009. The Type III Secretion System of Pseudomonas aeruginosa: Infection by Injection. Nature Reviews. *Microbiology*, 7 (9), pp.654–665.

Hirst T.R., Holmgren J., 1987. Conformation of protein secreted across bacterial outer membranes: a study of enterotoxin translocation from Vibrio cholerae. *Proc Natl Acad Sci* 84 (21), pp.7418-7422.

Ho, C.S., Chan M., Cheung R., Law L., Lit L., Ng K., Suen M., and Tai H., 2003. Electrospray Ionisation Mass Spectrometry: Principles and Clinical Applications. *Clin Biochem Re*v 24 (1), pp.3–12.

Hobot, J. A., E. Carlemalm, et al., 1984. Periplasmic gel: new concept resulting from the reinvestigation of bacterial cell envelope ultrastructure by new methods. *J Bacteriol.* 160 (1): 143-152.

Howard, S.P., 2013. Assembly of the Type II Secretion System. *Research in Microbiologoy*, 164 (6), pp.535–544.

Johnson, T.L., Scott, M.E. & Sandkvist, M., 2007. Mapping critical interactive sites within the periplasmic domain of the Vibrio cholerae type II secretion protein EpsM. *Journal of bacteriology*, 189 (24), pp.9082–9089.

Jones, H.E., Holland, I.B. & Campbell, A.K., 2002. Direct measurement of free Ca(2+) shows different regulation of Ca(2+) between the periplasm and the cytosol of Escherichia coli. *Cell calcium*, 32 (4), pp.183–192.

Jyot, J., Balloy, V., Jouvion, G., Verma, A., Touqui, L., Huerre, M., Ramphal, R., 2011. Type II Secretion System of Pseudomonas aeruginosa: In Vivo Evidence of a Significant Role in Death Due to Lung Infection. *The Journal of Infectious Diseases*, 203 (10), pp.1369–1377.

Kabsch, W. & Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22 (12), pp.2577–2637.

Karuppiah, V. et al., 2013. Structure and assembly of an inner membrane platform for initiation of type IV pilus biogenesis. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (48), pp.E4638–47.

Kelley, L.A. & Sternberg, M.J.E., 2009. Protein structure prediction on the Web: a case study using the Phyre server. *Nature protocols*, 4 (3), pp.363–371.

Klein, D.W., Prescott, LM.; Harley, John (2005). Microbiology. Boston: McGraw-Hill Higher Education. ISBN 0-07-295175-3.

Korotkov, K. V et al., 2009. Crystal structure of the N-terminal domain of the secretin GspD from ETEC determined with the assistance of a nanobody. *Structure* 17 (2), pp.255–265.

Korotkov, K. V et al., 2006. Structural and functional studies of EpsC, a crucial component of the type 2 secretion system from Vibrio cholerae. *Journal of molecular biology*, 363 (2), pp.311–321.

Korotkov, K. V et al., 2011. Structural and functional studies on the interaction of GspC and GspD in the type II secretion system. *PLoS pathogens*, 7(9), p.e1002228.

Korotkov, K. V, Sandkvist, M. & Hol, W.G.J., 2012. The type II secretion system: biogenesis, molecular architecture and mechanism. *Nature reviews. Microbiology*, 10 (5), pp.336–351.

Krissinel, E. & Henrick, K., 2004. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica. Section D, Biological crystallography*, 60 (Pt 12), pp.2256–2268.

Lallemand M, Login FH, Guschinskaya N, Pineau C, Effantin G, Robert X, Shevchik VE., 2013. Dynamic interplay between the periplasmic and transmembrane domains of GspL and GspM in the type II secretion system. PLoS One. 8 (11) pp.e79562

Langer G., Cohen S.X., Lamzin V.S., Perrakis A., 2008. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. *Nat Protoc.*, 3 (7) pp.1171-1179.

Larkin, M.A. et al., 2007. Clustal W and Clustal X version 2.0. *Bioinformatics (Oxford, England)*, 23 (21), pp.2947–2948.

Laskowski R.A., MacArthur M.W., Moss D.S., Thornton J.M., 1993. PROCHECK - a program to check the stereochemical quality of protein structures. *J. App. Cryst.*, 26, pp.283–291

Lindeberg M., Salmond G., Collmer A., 1996. Complementation of deletion mutations in a cloned functional cluster of Erwinia chrysanthemi out genes with Erwinia carotovora out homologues reveals OutC and OutD as candidate gatekeepers of species-specific secretion of proteins via the type II pathway. Mol Microbiol. 20 (1) pp.175–190.

Liu M., Mao X., Ye C., Huang H., Nicholson J., Lindon J., 1998. Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy, *Journal of Magnetic Resonance*, 132 (1), pp.125-129.

Login, F.H. & Shevchik, V.E., 2006. The single transmembrane segment drives self-assembly of OutC and the formation of a functional type II secretion system in Erwinia chrysanthemi. *The journal of biological chemistry*, 281 (44), pp.33152–33162.

Lybarger, S.R. et al., 2009. Docking and assembly of the type II secretion complex of Vibrio cholerae. *Journal of bacteriology*, 191 (9), pp.3149–3161.

Ma Q., Zhai Y., Schneider J.C., Ramseier T.M., Saier M.H., 2003. Protein secretion systems of Pseudomonas aeruginosa and P. fluorescens. *Biochim Biophys Acta*. 1611, pp.223-233.

Martin, S. R. and Schilstra, M. J., 2008. Circular Dichroism and Its Application to the Study of Biomolecules." *Methods in cell biology* 84: pp.263-293.

Martynowski, D. & Peter, S., 2013. Structure of a periplasmic domain of the EpsAB fusion protein of the Vibrio vulnificus type II secretion system research papers. *PLoS One*, pp.142–149.

McBride M. J. & Zhu Y., 2013. Gliding Motility and Por Secretion System Genes Are Widespread among members of the Phylum Bacteroidetes *Journal of Bacteriology* 195 (2), pp. 270 –278

McCoy, A. J. *et al.*, 2007. Phaser crystallographic software. *Journal of applied crystallography*, 40 (Pt 4), pp.658–674.

McLaughlin, L.S., Haft, R.J.F. & Forest, K.T., 2012. Structural insights into the Type II secretion nanomachine. *Current opinion in structural biology*, 22(2), pp.208–216.

Miroux B, Walker J., 1996. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol*, 260:289-98.

Misic, A.M., Satyshur, K.A. & Forest, K.T., 2010. P. aeruginosa PilT Structures with and without Nucleotide Reveal a Dynamic Type IV Pilus Retraction Motor. *Journal of Molecular Biology*, 400 (5), pp.1011–1021.

Muhlradt, P.F. & Golecki, J.R., 1975. Asymmetrical distribution and artifactual reorientation of lipopolysaccharide in the outer membrane bilayer of Salmonella typhimurium. *European journal of biochemistry / FEBS*, 51 (2), pp.343–352.

Nivaskumar, M. & Francetic, O., 2014. Biochimica et Biophysica Acta Type II secretion system : A magic beanstalk or a protein escalator. *BBA - Molecular Cell Research*, 1843 (8), pp.1568–1577.

Pauwels, K., Lustig A., Wyns L., Tommassen J., Savvides S.N., Van Gelder P., 2006. Structure of a membrane-based steric chaperone in complex with its lipase substrate. *Nature structural & molecular biology*, 13 (4), pp.374–375.

Palomäki T., Pickersgill R., Riekki R., Romantschuk M., Saarilahti H.T., 2002. A putative three-dimensional targeting motif of polygalacturonase (PehA), a protein secreted through the type II (GSP) pathway in Erwinia carotovora. *Mol Microbiol.* 43 (3), pp.585-596.

Peabody, C.R. et al., 2003. Type II protein secretion and its relationship to bacterial type IV pili and archaeal flagella. *Microbiology (Reading, England)*, 149 (Pt 11), pp.3051–3072.

Piddock, L., 2006. Multidrug-resistance efflux pumps? Not just for resistance. ***Nat Rev Micro*** **4 (8): 629-636.**

Pineau C, Guschinskaya N, Robert X, Gouet P, Ballut L, Shevchik VE., 2014. Substrate recognition by the bacterial type II secretion system: more than a simple interaction, ***Molecular Microbiology*****, 94, pp.126-140.**

Pugsley, A.P., Kornacker, M.G. & Poquet, I., 1991. The general protein-export pathway is directly required for extracellular pullulanase secretion in Escherichia coli K12. ***Molecular microbiology*****, 5 (2), pp.343–352.**

Pugsley A.P., Reyss I., 1990. Five genes at the 3′ end of the Klebsiella pneumoniae pulC operon are required for pullulanase secretion ***Molecular microbiology,*** **4 (3), pp. 365–379.**

Reichow, S.L. et al., 2010. Structure of the cholera toxin secretion channel in its closed state. ***Nature structural & molecular biology*****, 17 (10), pp.1226–1232.**

Reichow, S.L. et al., 2011. The binding of cholera toxin to the periplasmic vestibule of the type II secretion channel. ***Channels (Austin, Tex.)*****, 5 (3), pp.215–218.**

Robert, V. et al., 2002. Identification of XcpZ Domains Required for Assembly of the Secreton of Pseudomonas aeruginosa. ***Journal of Bacteriology*****, 184 (6), pp.1779–1782.**

Rondelet, A. & Condemine, G., 2013. Type II secretion : the substrates that won ' t go away. ***Research in Microbiologoy*****, 164 (6), pp.556–561.**

Rossier, O., Starkenburg, S.R. & Cianciotto, N.P., 2004. Legionella pneumophila type II protein secretion promotes virulence in the A/J mouse model of Legionnaires' disease pneumonia. *Infection and immunity*, 72 (1), pp.310–321.

Sandkvist, M. et al., 1995. Interaction between the autokinase EpsE and EpsL in the cytoplasmic membrane is required for extracellular secretion in Vibrio cholerae. *The EMBO journal*, 14 (8), pp.1664–1673.

Sandkvist, M., 2001. Type II secretion and pathogenesis. *Infection and immunity*, 69 (6), pp.3523–3535.

Schraidt O., Lefebre M.D., Brunner M.J., 2010. Topology and organization of the Salmonella typhimurium type III secretion needle complex components. *PLoS Pathogens*, 1 (6), pp.e1000824.

Shevchik VE1, Condemine G. 1998. Functional characterization of the Erwinia chrysanthemi OutS protein, an element of a type II secretion system. *Microbiology*, 144 (11), pp.3219–3228.

Shiue, S.J. et al., 2006. XpsE oligomerization triggered by ATP binding, not hydrolysis, leads to its association with XpsL. *The EMBO journal*, 25 (7), pp.1426–1435.

Strozen, T.G. et al., 2011. Involvement of the GspAB complex in assembly of the type II secretion system secretin of Aeromonas and Vibrio species. *Journal of bacteriology*, 193 (9), pp.2322–2331.

Thomas, J.D., Reeves, P.J. & Salmond, G.P.C., 1997. The general secretion pathway of Erwinia carotovora subsp. carotovora: analysis of the membrane topology of OutC and OutF. *Microbiology*, 143 (3), pp.713–720.

Vanderlinde, E.M. et al., 2014. Assembly of the type two secretion system in Aeromonas hydrophila involves direct interaction between the periplasmic domains of the assembly factor ExeB and the secretin ExeD. *PloS one*, 9 (7), p.e102038.

Walker, P.A. Leong L., Ng P., et al., 1994. Efficient and rapid affinity purification of proteins using recombinant fusion proteases. *Biotechnology* 12 (6), pp.601-605.

Wang, X. et al., 2012. Cysteine scanning mutagenesis and disulfide mapping analysis of arrangement of GspC and GspD protomers within the type 2 secretion system. *The Journal of biological chemistry*, 287 (23), pp.19082–19093.

Weiss, A. A., Johnson, F. D. & Burns, D. L., 1993. Molecular characterization of an operon required for pertussis toxin secretion. *Proc. Natl Acad. Sci. USA*, 90, pp. 2970–2974.

Winn M.D., Murshudov G.N., Papiz M.Z., 2003. Macromolecular TLS refinement in REFMAC at moderate resolutions. *Methods in Enzymology*, 374, pp.300-321.

Winter, G., Lobley, C.M.C. & Prince, S.M., 2013. Decision making in xia2. *Acta crystallographica. Section D, Biological crystallography*, 69 (7), pp.1260–1273.

Yamashita, M., Fenn., 1984. Electrospray ion source. Another variation on the free-jet theme. *Journal of Physical Chemistry*, 88 (20), pp.4451–4459.

Yang S., Peng Q., Francisco M., Wang Y., Zeng Q., et al., 2008. Type III Secretion System Genes of Dickeya dadantii 3937 Are Induced by Plant Phenolic Acids. *PLoS ONE*, 3 (8) pp.e2973

Zechner E., Lang S., and Schildbach J., 2012. Assembly and mechanisms of bacterial type IV secretion machines *Philos Trans R Soc Lond B Biol Sci.* 367 (1592), pp.1073–1087.